(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 August 2001 (09.08.2001)

PCT

(10) International Publication Number
WO 01/57275 A2

(51) International Patent Classification[7]: C12Q 1/68, G06F 19/00, C07K 14/47

(21) International Application Number: PCT/US01/00667

(22) International Filing Date: 30 January 2001 (30.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/180,312	4 February 2000 (04.02.2000)	US
60/207,456	26 May 2000 (26.05.2000)	US
09/608,408	30 June 2000 (30.06.2000)	US
09/632,366	3 August 2000 (03.08.2000)	US
60/234,687	21 September 2000 (21.09.2000)	US
60/236,359	27 September 2000 (27.09.2000)	US
0024263.6	4 October 2000 (04.10.2000)	GB

(71) Applicant *(for all designated States except US)*: MOLECULAR DYNAMICS, INC. [—/US]; 928 East Arques Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PENN, Sharron, G. [GB/US]; 617 South Delaware Street, San Mateo, CA 94402 (US). HANZEL, David, K. [US/US]; 988 Loma Verde Avenue, Palo Alto, CA 94303 (US). CHEN, Wensheng [CN/US]; 210 Easy Street #25, Mountain View, CA 94043 (US). RANK, David, R. [US/US]; 117 El Dorado Commons, Fremont, CA 94539 (US).

(74) Agent: RONNING, Royal, N., Jr.; Amersham Pharmacia Biotech, Inc., 800 Centennial Avenue, Piscataway, NJ 08855 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/57275 A2

(54) Title: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR ANALYSIS OF GENE EXPRESSION IN HUMAN BRAIN

(57) Abstract: A single exon nucleic acid microarray comprising a plurality of single exon nucleic acid probes for measuring gene expression in a sample derived from human brain is described. Also described are single exon nucleic acid probes expressed in the brain and their use in methods for detecting gene expression.

AF2

HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR ANALYSIS OF GENE EXPRESSION IN HUMAN BRAIN

CROSS REFERENCE TO RELATED APPLICATIONS

The present application is a continuation-in-part of U.S. patent application serial nos. 09/632,366, filed August 3, 2000 and 09/608,408, filed June 30, 2000; claims the benefit under 35 U.S.C. s 119(e) of U.S.provisional patent application serial nos. 60/236,359, filed September 27, 2000, 60/234,687, filed September 21, 2000, 60/207,456, filed May 26, 2000, and 60/180,312, filed February 4, 2000; and further claims the benefit under 35 U.S.C. s 119(a) of UK patent application no. 0024263.6, filed October 4, 2000, the disclosures of which are incorporated herein by reference in their entireties.

REFERENCE TO SEQUENCE LISTING AND INCORPORATION BY REFERENCE THEREOF

The present application includes a Sequence Listing in electronic format, filed pursuant to PCT Administrative Instructions 801 - 806 on a single CD-R disc, in triplicate, containing a file named pto_BRAIN.txt, created 24 January 2001, having 25,840,972 bytes. The Sequence Listing contained in said file on said disc is incorporated herein by reference in its entirety.

Field of the Invention

The present invention relates to genome-derived single exon microarrays useful for verifying the expression of regions of genomic DNA predicted to encode protein. In particular, the present invention relates to unique genome-derived single exon nucleic acid probes expressed in human

brain and single exon nucleic acid microarrays that include such probes.

Background of the Invention

For almost two decades following the invention of general techniques for nucleic acid sequencing, Sanger *et al.*, *Proc. Natl. Acad. Sci. USA* 70(4):1209-13 (1973); Gilbert *et al.*, *Proc. Natl. Acad. Sci. USA* 70(12):3581-4 (1973), these techniques were used principally as tools to further the understanding of proteins – known or suspected – about which a basic foundation of biological knowledge had already been built. In many cases, the cloning effort that preceded sequence identification had been both informed and directed by that antecedent biological understanding.

For example, the cloning of the T cell receptor for antigen was predicated upon its known or suspected cell type-specific expression, by its suspected membrane association, and by the predicted assembly of its gene via T cell-specific somatic recombination. Subsequent sequencing efforts at once confirmed and extended understanding of this family of proteins. Hedrick *et al.*, *Nature* 308(5955):153-8 (1984).

More recently, however, the development of high throughput sequencing methods and devices, in concert with large public and private undertakings to sequence the human and other genomes, has altered this investigational paradigm: today, sequence information often precedes understanding of the basic biology of the encoded protein product.

One of the approaches to large-scale sequencing is predicated upon the proposition that expressed sequences – that is, those accessible through isolation of mRNA – are of greatest initial interest. This "expressed sequence tag" ("EST") approach has already yielded vast

amounts of sequence data (see for example Adams *et al.*, *Science* 252:1651 (1991); Williamson, *Drug Discov. Today* 4:115 (1999)). For nucleic acids sequenced by this approach, often the only biological information that is known *a priori* with any certainty is the likelihood of biologic expression itself. By virtue of the species and tissue from which the mRNA had originally been obtained, most such sequences are also annotated with the identity of the species and at least one tissue in which expression appears likely.

More recently, the pace of genomic sequencing has accelerated dramatically. When genomic DNA serves as the initial substrate for sequencing efforts, expression cannot be presumed; often the only *a priori* biological information about the sequence includes the species and chromosome (and perhaps chromosomal map location) of origin.

With the ever-accelerating pace of sequence accumulation by directed, EST, and genomic sequencing approaches – and in particular, with the accumulation of sequence information from multiple genera, from multiple species within genera, and from multiple individuals within a species – there is an increasing need for methods that rapidly and effectively permit the functions of nucleic sequences to be elucidated. And as such functional information accumulates, there is a further need for methods of storing such functional information in meaningful and useful relationship to the sequence itself; that is, there is an increasing need for means and apparatus for annotating raw sequence data with known or predicted functional information.

Although the increase in the pace of genomic sequencing is due in large part to technological changes in sequencing strategies and instrumentation, Service, *Science* 280:995 (1998); Pennisi, *Science* 283: 1822-1823 (1999), there is an important functional motivation as well.

While it was understood that the EST approach would rarely be able to yield sequence information about the noncoding portions of the genome, it now also appears the EST approach is capable of capturing only a fraction of a genome's actual expression complexity.

For example, when the *C. elegans* genome was fully sequenced, gene prediction algorithms identified over 19,000 potential genes, of which only 7,000 had been found by EST sequencing. *C. elegans* Sequencing Consortium, *Science* 282:2012 (1998). Analogously, the recently completed sequence of chromosome 2 of *Arabidopsis* predicts over 4000 genes, Lin *et al.*, *Nature*, 402:761 (1999), of which only about 6% had previously been identified via EST sequencing efforts. Although the human genome has the greatest depth of EST coverage, it is still woefully short of surrendering all of its genes. One recent estimate suggests that the human genome contains more than 146,000 genes, which would at this point leave greater than half of the genes undiscovered. It is now predicted that many genes, perhaps 20 to 50%, will only be found by genomic sequencing.

There is, therefore, a need for methods that permit the functional regions of genomic sequence – and most importantly, but not exclusively, regions that function to encode genes – to be identified.

Much of the coding sequence of the human genome is not homologous to known genes, making detection of open reading frames ("ORFs") and predictions of gene function difficult. Computational methods exist for predicting coding regions in eukaryotic genomes. Gene prediction programs such as GRAIL and GRAIL II, Uberbacher *et al.*, *Proc. Natl. Acad. Sci. USA* 88(24):11261-5 (1991); Xu *et al.*, *Genet. Eng.* 16:241-53 (1994); Uberbacher *et al.*, *Methods Enzymol.* 266:259-81 (1996); GENEFINDER, Solovyev *et al.*, *Nucl. Acids. Res.* 22:5156-63 (1994); Solovyev *et al.*,

Ismb 5:294-302 (1997); and GENESCAN, Burge *et al.*, *J. Mol. Biol.* 268:78-94 (1997), predict many putative genes without known homology or function. Such programs are known, however, to give high false positive rates. Burset *et al.*, *Genomics* 34:353-367 (1996). Using a consensus obtained by a plurality of such programs is known to increase the reliability of calling exons from genomic sequence. Ansari-Lari *et al.*, *Genome Res.* 8(1):29-40 (1998)

Identification of functional genes from genomic data remains, however, an imperfect art. For example, in reporting the full sequence of human chromosome 21, the Chromosome 21 Mapping and Sequencing Consortium reports that prior bioinformatic estimates of human gene number may need to be revised substantially downwards. *Nature* 405:311-199 (2000); Reeves, *Nature* 405:283-284 (2000).

Thus, there is a need for methods and apparatus that permit the functions of the regions identified bioinformatically – and specifically, that permit the expression of regions predicted to encode protein – readily to be confirmed experimentally.

Recently, the development of nucleic acid microarrays has made possible the automated and highly parallel measurement of gene expression. *Reviewed in* Schena (ed.), <u>DNA Microarrays : A Practical Approach (Practical Approach Series)</u>, Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1 - 60 (1999); Schena (ed.), <u>Microarray Biochip: Tools and Technology</u>, Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376).

It is common for microarrays to be derived from cDNA/EST libraries, either from those previously described in the literature, such as those from the I.M.A.G.E. consortium, Lennon *et al.*, *Genomics* 33(1):151-2 (1996), or from the construction of "problem specific" libraries targeted at a particular biological question, R.S. Thomas

et al., *Cancer Res.* (in press). Such microarrays by definition can measure expression only of those genes found in EST libraries, and thus have not been useful as probes for genes discovered solely by genomic sequencing.

The utility of using whole genome nucleic acid microarrays to answer certain biological questions has been demonstrated for the yeast *Saccharomyces cerevisiae*. De Risi *et al.*, *Science* 278:680 (1997). The vast majority of yeast nuclear genes, approximately 95% however, are single exon genes, *i.e.*, lack introns, Lopez *et al.*, *RNA* 5:1135-1137 (1999); Goffeau *et al.*, *Science* 274:563-67 (1996), permitting coding regions more readily to be identified. Whole genome nucleic acid microarrays have not generally been used to probe gene expression from more complex eukaryotic genomes, and in particular from those averaging more than one intron per gene.

Diseases of the brain and nervous system are a significant cause of human morbidity and mortality. Increasingly, genetic factors are being found that contribute to predisposition, onset, and/or aggressiveness of most, if not all, of these diseases. Although mutations in single genes have been identified as causative for some diseases of the brain and nervous system, for the most part these disorders are believed to have polygenic etiologies. There is a need for methods and apparatus that permit prediction, diagnosis and prognosis of diseases of the brain and nervous system particularly those diseases with polygenic etiology.

Summary of the Invention

The present invention solves these and other problems in the art by providing methods and apparatus for predicting, confirming, and displaying functional information derived from genomic sequence. The present

invention also provides apparatus for verifying the expression of putative genes identified within genomic sequence.

In particular, the invention provides novel genome-derived single exon nucleic acid microarrays useful for verifying the expression of putative genes identified within genomic sequence.

The present invention also provides compositions and kits for the ready production of nucleic acids identical in sequence to, or substantially identical in sequence to, probes on the genome-derived single exon microarrays of the present invention.

Accordingly, in a first aspect of the invention, there is provided a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human brain, comprising a plurality of single exon nucleic acid probes according to any one of the nucleotide sequences set out in SEQ ID NOs: 1 - 12,821 or a complementary sequence, or a portion of such a sequence.

By plurality is meant at least two, suitably at least 20, most suitably at least 100, preferably at least 1000 and, most preferably, upto 5000.

In one embodiment of the first aspect, each of said plurality of probes is separately and addressably amplifiable.

In an alternative embodiment, each of said plurality of probes is separately and addressably isolatable from said plurality.

In a preferred embodiment, each of said plurality of probes is amplifiable using at least one common primer. Preferably, each of said plurality of probes is amplifiable using a first and a second common primer.

In yet another embodiment, said set of single exon nucleic acid probes comprises between 50 - 20,000 probes, for example, 50 - 5000.

Suitably, said set of single exon nucleic acid probes comprises at least 50 - 1000 discrete single exon nucleic acid probes having a sequence as set out in any of SEQ ID NOS.: 1 - 25,434 or a complimentary sequence, or a portion of such a sequence.

Preferably, the average length of the single exon nucleic acid probes is between 200 and 500 bp. It is preferred that the average length should be at least 200bp, suitably at least 250bp, most suitably at least 300bp, preferably at least 400bp and, most preferably, 500 bp.

In another embodiment, the single exon nucleic acid probes lack prokaryotic and bacteriophage vector sequence. It is preferred that at least 50%, suitably at least 60%, most suitably at least 70%, preferably at least 75%, more preferably at least 80, 85, 90, 95 or 99% of said single exon nucleic acid probes lack prokaryotic and bacteriophage vector sequence.

In another preferred embodiment, said single exon nucleic acid lack homopolymeric stretches of A or T. It is preferred that at least 50%, suitably at least 60%, most suitably at least 70%, preferably at least 75%, more preferably at least 80, 85, 90, 95 or 99% of said single exon nucleic acid probes lack homopolymeric stretches of A or T.

Preferably, a spatially-addressable set of single exon nucleic acid probes in accordance with the first aspect of the invention is is addressably disposed upon a substrate.

Suitable substrates include a filter membrane which may, preferably, be nitrocellulose or nylon. The nylon may preferably, be positively-charged. Other suitable substrates include glass, amorphous silicon, crystalline silicon, and plastic. Further suitable materials include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride,

polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, and mixtures thereof.

In a second aspect of the invention, there is provided a microarray comprising a spatially addressable set of single exon nucleic acid probes in accordance with the first aspect of the invention.

In one embodiment, a genome-derived single-exon microarray is packaged together with such an ordered set of amplifiable probes corresponding to the probes, or one or more subsets of probes, thereon. In alternative embodiments, the ordered set of amplifiable probes is packaged separately from the genome-derived single exon microarray.

In another aspect, the invention provides genome-derived single exon nucleic acid probes useful for gene expression analysis, and particularly for gene expression analysis by microarray. In particular embodiments of this aspect, the present invention provides human single-exon probes that include specifically-hybridizable fragments of SEQ ID Nos. 12,822 - 25,434, wherein the fragment hybridizes at high stringency to an expressed human gene. In particular embodiments, the invention provides single exon probes comprising SEQ ID Nos. 1 - 12,821.

Accordingly, in a third aspect of the invention, there is provided a single exon nucleic acid probe for measuring human gene expression in a sample derived from human brain which is a nucleic acid molecule comprising a nucleotide sequence as set out in any of SEQ ID NOs.: 1 - 12,821 or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid expressed in the human brain.

In one embodiment, a single exon nucleic acid probe in accordance with the third aspect comprises a nucleotide sequence as set out in any of SEQ ID NOs.:

12,822 - 25,434 or a complementary sequence or a fragment thereof.

In a fourth aspect of the invention, there is provided a single exon nucleic acid probe for measuring human gene expression in a sample derived from human brain which is a nucleic acid molecule having a sequence encoding a peptide comprising a peptide sequence as set out in any of SEQ ID NOs.: 25,435 - 37,811or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid expressed in the human brain.

Preferably, a single exon nucleic acid probe in accordance with the third or fourth aspects of the invention comprises between at least 15 and 50 contiguous nucleotides of said SEQ ID NO:. It is preferred that the single exon nucleic acid probe comprises at least 15, suitably at least 20, more suitably at least 25 or preferably at least 50 contiguous nucleotides of said SEQ ID NO:.

In another preferred embodiment, a single exon nucleic acid probe in accordance with the third or fourth aspects of the invention is between 3kb and 25kb in length. It is preferred that said probe is no more than 3kb, suitably no more than 5kb, more suitably no more than 10kb, preferably 15kb, more preferably 20kb or, most preferably, no more than 20kb in length.

Preferably, a single exon nucleic acid probe in accordance with either the fifth or sixth aspect of the invention is DNA, preferably single-stranded DNA, RNA or PNA.

In another embodiment of either the third or fourth aspect of the invention, a single exon nucleic acid probe is detectably labeled. Suitable detectable labels include a radionuclide, a fluorescent label or a first member of a specific binding pair. Suitable fluorescent labels include dyes such as cyanine dyes, preferably Cy3

and Cy5 although other suitable dyes will be known to those skilled in the art.

In a particularly preferred embodiment, a single exon nucleic acid probe in accordance with either the third or fourth aspect of the invention lacks prokaryotic and bacteriophage vector sequence. In yet another embodiment, a single exon nucleic acid probe in accordance with either the third or fourth aspect of the invention lacks homopolymeric stretches of A or T.

In a fifth aspect of the invention, there is provided an amplifiable nucleic acid composition, comprising:

the single exon nucleic acid probe in accordance with either of the third or fourth aspects of the invention; and at least one nucleic acid primer;

wherein said at least one primer is sufficient to prime enzymatic amplification of said probe.

In an sixth aspect of the invention, there is provided a method of measuring gene expression in a sample derived from human brain, comprising:

contacting the single exon microarray in accordance with the second aspect of the invention, with a first collection of detectably labeled nucleic acids, said first collection of nucleic acids derived from mRNA of human brain; and then

measuring the label detectably bound to each probe of said microarray.

In a seventh aspect of the invention, there is provided a method of identifying exons in a eukaryotic genome, comprising:

algorithmically predicting at least one exon from genomic sequence of said eukaryote; and then

detecting specific hybridization of detectably labeled nucleic acids to a single exon probe,

wherein said detectably labeled nucleic acids are

derived from mRNA from the brain of said eukaryote, said probe is a single exon probe having a fragment identical in sequence to, or complementary in sequence to, said predicted exon, said probe is included within a single exon microarray in accordance with the first aspect of the invention, and said fragment is selectively hybridizable at high stringency.

In a eighth aspect of the invention, there is provided a method of assigning exons to a single gene, comprising:

identifying a plurality of exons from genomic sequence in accordance with the seventh aspect of the invention; and then

measuring the expression of each of said exons in a plurality of tissues and/or cell types using hybridization to single exon microarrays having a probe with said exon,

wherein a common pattern of expression of said exons in said plurality of tissues and/or cell types indicates that the exons should be assigned to a single gene.

In an ninth aspect of the invention, there is provided a nucleic acid sequence as set out in any of SEQ ID NOs: 1 - 25,434 wherein said sequence encodes a peptide.

In a tenth aspect of the invention, there is provided a peptide encoded by a sequence comprising a sequence as set out in any of SEQ ID NOs: 12,822 - 25,434, or a complementary sequence or coding portion thereof.

In a preferred embodiment, a peptide may be encoded by a sequence comprising a sequence set out in any of SEQ ID NOS.: 1 -12,821.

In a further aspect, the invention provides peptides comprising an amino acid sequence translated from the DNA fragments, said amino acid sequences comprising SEQ ID NOS.: 25,435 - 37,811.

Accordingly in a eleventh aspect of the invention there is provided a peptide comprising a sequence as set out in any of SEQ ID NOs: 25,435 - 37,811, or fragment thereof.

In another aspect, the invention provides means for displaying annotated sequence, and in particular, for displaying sequence annotated according to the methods and apparatus of the present invention. Further, such display can be used as a preferred graphical user interface for electronic search, query, and analysis of such annotated sequence.

Detailed Description of the Invention

Definitions

As used herein, the term "microarray" and phrase "nucleic acid microarray" refer to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed.

As so defined, the term "microarray" and phrase "nucleic acid microarray" include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1 - 60 (1999); and Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376). As so defined, the term "microarray" and phrase "nucleic acid microarray" further include substrate-bound collections of plural nucleic acids in which the nucleic acids are distributably disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner

et al., *Proc. Natl. Acad. Sci. USA* 97(4):166501670 (2000); in such case, the term "microarray" and phrase "nucleic acid microarray" refer to the plurality of beads in aggregate.

As used herein with respect to a nucleic acid microarray, the term "probe" refers to the nucleic acid that is, or is intended to be, bound to the substrate; in such context, the term "target" thus refers to nucleic acid intended to be bound thereto by Watson-Crick complementarity. As used herein with respect to solution phase hybridization, the term "probe" refers to the nucleic acid of known sequence that is detectably labeled.

As used herein, the expression "probe comprising SEQ ID NO.", and variants thereof, intends a nucleic acid probe, at least a portion of which probe has either (i) the sequence directly as given in the referenced SEQ ID NO., or (ii) a sequence complementary to the sequence as given in the referenced SEQ ID NO., the choice as between sequence directly as given and complement thereof dictated by the requirement that the probe hybridize to mRNA.

As used herein, the term "open reading frame" and the equivalent acronym "ORF" refer to that portion of an exon that can be translated in its entirety into a sequence of contiguous amino acids i.e. a nucleic acid sequence that, in at least one reading frame, does not possess stop codons; the term does not require that the ORF encode the entirety of a natural protein.

As used herein, the term "amplicon" refers to a PCR product amplified from human genomic DNA, containing the predicted exon.

As used herein the term "exon" refers to the consensus prediction of the various exon and gene predicting algorithms i.e. a nucleic acid sequence bioinformatically predicted to encode a portion of a natural protein.

As used herein, the term "peptide" refers to a sequence of amino acids. The sequences referred to as PEPTIDE SEQ ID NOS.: are the predicted peptide sequences that would be translated from one of the exons, or a portion thereof set out in exon SEQ ID NOS.:. The codons encoding the peptide are wholly contained within the exon.

As used herein, a "portions" of a defined nucleotide sequence or sequences can be and, preferably, are fragments unique to that sequence or to one or a combination of those sequences. A fragment unique to a nucleic acid molecule is one that is a signature for the larger nucleic acid molecule.

As used herein, the phrase "expression of a probe" and its linguistic variants means that the ORF present within the probe, or its complement, is present within a target mRNA.

As used herein, "stringent conditions" refers to parameters well known to those skilled in the art. When a nucleic acid molecule is said to be hybridisable to another of a given sequence under "stringent conditions" it is meant that it is homologous to the given sequence.

As used herein, the phrase "specific binding pair" intends a pair of molecules that bind to one another with high specificity. Binding pairs are said to exhibit specific binding when they exhibit avidity of at least 10^7, preferably at least 10^8, more preferably at least 10^9 liters/mole. Nonlimiting examples of specific binding pairs are: antibody and antigen; biotin and avidin; and biotin and streptavidin.

As used herein with respect to the visual display of annotated genomic sequence, the term "rectangle" means any geometric shape that has at least a first and a second border, wherein the first and second borders each are capable of mapping uniquely to a point of another visual object of the display.

As used herein, a "Mondrian" means a visual display in which a single genomic sequence is annotated with predicted and experimentally confirmed functional information.

Brief Description of the Drawings

The present invention is further illustrated with reference to the following non-limiting figures and examples in which:

FIG. 1 illustrates a process for predicting functional regions from genomic sequence, confirming the functional activity of such regions experimentally, and associating and displaying the data so obtained in meaningful and useful relationship to the original sequence data;

FIG. 2 further elaborates that portion of the process schematized in FIG. 1 for predicting functional regions from genomic sequence;

FIG. 3 illustrates a Mondrian visual display;

FIG. 4 presents a Mondrian showing a hypothetical annotated genomic sequence;

FIG. 5 is a histogram showing the distribution of ORF length and PCR products as obtained, with ORF length shown in black and PCR product length shown in dotted lines;

FIG. 6 is a histogram showing the distribution, among exons predicted according to the methods described, of expression as measured using simultaneous two color hybridization to a genome-derived single exon microarray. The graph shows the number of sequence-verified products that were either not expressed ("0"), expressed in one or more but not all tested tissues ("1" - "9"), or expressed in all tissues tested ("10");

FIG. 7 is a pictorial representation of the expression of verified sequences that showed expression with signal intensity greater than 3 in at least one tissue, with: FIG. 7A showing the expression as measured by microarray hybridization in each of the 10 measured tissues, and the expression as measured "bioinformatically" by query of EST, NR and SwissProt databases; with FIG. 7B showing the legend for display of physical expression (ratio) in FIG. 7A; and with FIG. 7C showing the legend for scoring EST hits as depicted in FIG. 7A;

FIG. 8 shows a comparison of normalized CY3 signal intensity for arrayed sequences that were identical to sequences in existing EST, NR and SwissProt databases or that were dissimilar (unknown), where black denotes the signal intensity for all sequence-verified products with a BLAST Expect ("E") value of greater than 1e-30 (1×10^{-30}) ("unknown") and a dotted line denotes sequence-verified spots with a BLAST expect ("E") value of less than 1e-30 (1×10^{-30}) ("known");

FIG. 9 presents a Mondrian of BAC AC008172 (bases 25,000 to 130,000), containing the carbamyl phosphate synthetase gene (AF154830.1); and

FIG. 10 is a Mondrian of BAC A049839.

Methods and Apparatus for Predicting, Confirming, Annotating, and Displaying Functional Regions From Genomic Sequence Data

FIG. 1 is a flow chart illustrating in broad outline a process for predicting functional regions from genomic sequence, confirming and characterizing the functional activity of such regions experimentally, and then associating and displaying the information so obtained in meaningful and useful relationship to the original

sequence data.

The initial input into process 10 of the present invention is drawn from one or more databases 100 containing genomic sequence data. Because genomic sequence is usually obtained from subgenomic fragments, the sequence data typically will be stored in a series of records corresponding to these subgenomic sequenced fragments. Some fragments will have been catenated to form larger contiguous sequences ("contigs"); others will not. A finite percentage of sequence data in the database will typically be erroneous, consisting *inter alia* of vector sequence, sequence created from aberrant cloning events, sequence of artificial polylinkers, and sequence that was erroneously read.

Each sequence record in database 100 will minimally contain as annotation a unique sequence identifier (accession number), and will typically be annotated further to identify the date of accession, species of origin, and depositor. Because database 100 can contain nongenomic sequence, each sequence will typically be annotated further to permit query for genomic sequence. Chromosomal origin, optionally with map location, can also be present. Data can be, and over time increasingly will be, further annotated with additional information, in part through use of the present invention, as described below. Annotation can be present within the data records, in information external to database 100 and linked to the records thereto, or through a combination of the two.

Databases useful as genomic sequence database 100 in the present invention include GenBank, and particularly include several divisions thereof, including the htgs(draft), NT (nucleotide, command line), and NR (nonredundant) divisions. GenBank is produced by the National Institutes of Health and is maintained by the National Center for Biotechnology Information (NCBI).

Databases of genomic sequence from species other than human, such as mouse, rat, Arabidopsis, *C. elegans*, *C. brigsii*, *Drosophila*, zebra fish, and other higher eukaryotic organisms will also prove useful as genomic sequence database 100.

Genomic sequence obtained by query of genomic sequence database 100 is then input into one or more processes 200 for identification of regions therein that are predicted to have a biological function as specified by the user. Such functions include, but are not limited to, encoding protein, regulating transcription, regulating message transport after transcription into mRNA, regulating message splicing after transcription into mRNA, of regulating message degradation after transcription into mRNA, and the like. Other functions include directing somatic recombination events, contributing to chromosomal stability or movement, contributing to allelic exclusion or X chromosome inactivation, and the like.

The particular genomic sequence to be input into process 200 will depend upon the function for which relevant sequence is to be identified as well as upon the approach chosen for such identification. Process step 200 can be iterated to identify different functions within a given genomic region. In such case, the input often will be different for the several iterations.

Sequences predicted to have the requisite function by process 200 are then input into process 300, where a subset of the input sequences suitable for experimental confirmation is identified. Experimental confirmation can involve physical and/or bioinformatic assay. Where the subsequent experimental assay is bioinformatic, rather than physical, there are fewer constraints on the sequences that can be tested, and in this latter case therefore process 300 can output the entirety of the input sequence.

The subset of sequences output from process 300 is then used in process 400 for experimental verification and characterization of the function predicted in process 200, which experimental verification can, and often will, include both physical and bioinformatic assay.

Process 500 annotates the sequence data with the functional information obtained in the physical and/or bioinformatic assays of process 400. Such annotation can be done using any technique that usefully relates the functional information to the sequence, as, for example, by incorporating the functional data into the sequence data record itself, by linking records in a hierarchical or relational database, by linking to external databases, by a combination thereof, or by other means well known within the database arts. The data can even be submitted for incorporation into databases maintained by others, such as GenBank, which is maintained by NCBI.

As further noted in FIG. 1, additional annotation can be input into process 500 from external sources 600.

The annotated data is then displayed in process 800, either before, concomitantly with, or after optional storage 700 on nontransient media, such as magnetic disk, optical disc, magnetooptical disk, flash memory, or the like.

FIG. 1 shows that the experimental data output from process 400 can be used in each preceding step of process 10: e.g., facilitating identification of functional sequences in process 200, facilitating identification of an experimentally suitable subset thereof in process 300, and facilitating creation of physical and/or informational substrates for, and performance of subsequent assay, of functional sequences in process 400.

Information from each step can be passed directly to the succeeding process, or stored in permanent or interim form prior to passage to the succeeding process.

Often, data will be stored after each, or at least a plurality, of such process steps. Any or all process steps can be automated.

FIG. 2 further elaborates the prediction of functional sequence within genomic sequence according to process 200.

Genomic sequence database 100 is first queried 20 for genomic sequence.

The sequence required to be returned by query 20 will depend, in the first instance, upon the function to be identified.

For example, genomic sequences that function to encode protein can be identified *inter alia* using gene prediction approaches, comparative sequence analysis approaches, or combinations of the two. In gene prediction analysis, sequence from one genome is input into process 200 where at least one, preferably a plurality, of algorithmic methods are applied to identify putative coding regions. In comparative sequence analysis, by contrast, corresponding, *e.g.*, syntenic, sequence from a plurality of sources, typically a plurality of species, is input into process 200, where at least one, possibly a plurality, of algorithmic methods are applied to compare the sequences and identify regions of least variability.

The exact content of query 20 will also depend upon the database queried. For example, if the database contains both genomic and nongenomic sequence, perhaps derived from multiple species, and the function to be determined is protein coding regions in human genomic sequence, the query will accordingly require that the sequence returned be genomic and derived from humans.

Query 20 can also incorporate criteria that compel return of sequence that meets operative requirements of the subsequent analytical method. Alternatively, or in addition, such operative criteria can be enforced in

subsequent preprocess step 24.

For example, if the function sought to be identified is protein coding, query 20 can incorporate criteria that return from genomic sequence database 100 only those sequences present within contigs sufficiently long as to have obviated substantial fragmentation of any given exon among a plurality of separate sequence fragments.

Such criteria can, for example, consist of a required minimal individual genomic sequence fragment length, such as 10 kb, more typically 20 kb, 30 kb, 40kb, and preferably 50 kb or more, as well as an optional further or alternative requirement that sequence from any given clone, such as a bacterial artificial chromosome ("BAC"), be presented in no more than a finite maximal number of fragments, such as no more than 20 separate pieces, more typically no more than 15 fragments, even more typically no more than about 10 - 12 fragments.

Results using the present invention have shown that genomic sequence from bacterial artificial chromosomes (BACs) is sufficient for gene prediction analysis according to the present invention if the sequence is at least 50 kb in length, and if additionally the sequence from any given BAC is presented in fewer than 15, and preferably fewer than 10, fragments. Accordingly, query 20 can incorporate a requirement that data accessioned from BAC sequencing be in fewer than 15, preferably fewer than 10, fragments.

An additional criterion that can be incorporated into the query can be the date, or range of dates, of sequence accession. Although the process has been described above as if genomic sequence database 100 were static, it is of course understood that the genomic sequence databases need not be static, and indeed are typically updated on a frequent, even hourly, basis. Thus, as further described in Examples 1 and 2, *infra*, it is

possible to query the database for newly added sequence, either newly added after an absolute date, or newly added relative to a prior analysis performed using the methods and apparatus of the present invention. In this way, the process herein described can incorporate a dynamic, temporal component.

One utility of such temporal limitation is to identify, from newly accessioned genomic sequence, the presence of novel genes, particularly those not previously identified by EST sequencing (or other sequencing efforts that are similarly based upon gene expression). As further described in Example 1, such an approach has shown that newly accessioned human genomic sequence, when analyzed for sequences that function to encode protein, readily identifies genes that are novel over those in existing EST and other expression databases. This makes the methods of the present invention extremely powerful gene discovery tools. And as would be appreciated, such gene discovery can be performed using genomic sequence from species other than human.

If query 20 incorporates multiple criteria, such as above-described, the multiple criteria can be performed as a series of separate queries or as a single query, depending in part upon the query language, the complexity of the query, and other considerations well known in the database arts.

If query 20 returns no genomic sequence meeting the query criteria, the negative result can be reported by process 22, and process 200 (and indeed, entire process 10) ended 23, as shown. Alternatively, or in addition to report and termination of the initial inquiry, a new query 20 can be generated that takes into account the initial negative result.

When query 20 returns sequence meeting the query criteria, the returned sequence is then passed to optional

preprocessing 24, suitable and specific for the desired analytical approach and the particular analytical methods thereof to be used in process 25.

Preprocessing 24 can include processes suitable for many approaches and methods thereof, as well as processes specifically suited for the intended subsequent analysis.

Preprocessing 24 suitable for most approaches and methods will include elimination of sequence irrelevant to, or that would interfere with, the subsequent analysis. Such sequence includes repetitive sequence, such as Alu repeats and LINE elements, vector sequence, artificial sequence, such as artificial polylinkers, and the like. Such removal can readily be performed by identification and subsequent masking of the undesired sequence.

Identification can be effected by comparing the genomic sequence returned by query 20 with public or private databases containing known repetitive sequence, vector sequence, artificial sequence, and other artifactual sequence. Such comparison can readily be done using programs well known in the art, such as CROSS_MATCH, or by proprietary sequence comparison programs the engineering of which is well within the skill in the art.

Alternatively, or in addition, undesirable, including artifactual, sequence can be identified algorithmically without comparison to external databases and thereafter removed. For example, synthetic polylinker sequence can be identified by an algorithm that identifies a significantly higher than average density of known restriction sites. As another example, vector sequence can be identified by algorithms that identify nucleotide or codon usage at variance with that of the bulk of the genomic sequence.

Once identified, undesired sequence can be removed. Removal can usefully be done by masking the

undesired sequence as, for example, by converting the specific nucleotide references to one that is unrecognized by the subsequent bioinformatic algorithms, such as "X". Alternatively, but at present less preferred, the undesired sequence can be excised from the returned genomic sequence, leaving gaps.

Preprocessing 24 can further include selection from among duplicative sequences of that one sequence of highest quality. Higher quality can be measured as a lower percentage of, fewest number of, or least densely clustered occurrence of ambiguous nucleotides, defined as those nucleotides that are identified in the genomic sequence using symbols indicating ambiguity. Higher quality can also or alternatively be valued by presence in the longest contig.

Preprocessing 24 can, and often will, also include formatting of the data as specifically appropriate for passage to the analytical algorithms of process 25. Such formatting can and typically will include, *inter alia*, addition of a unique sequence identifier, either derived from the original accession number in genomic sequence database 100, or newly applied, and can further include additional annotation. Formatting can include conversion from one to another sequence listing standard, such as conversion to or from FASTA or the like, depending upon the input expected by the subsequent process.

Preprocessing, which can be optional depending upon the function desired to be identified and the informational requirements of the methods for effecting such identification, is followed by sequence processing 25, where sequences with the desired function are identified within the genomic sequence.

As mentioned above, such functions can include, but are not limited to, encoding protein, regulating transcription, regulating message transport after

transcription into mRNA, regulating message splicing after transcription, of regulating message degradation, and the like. Other functions include directing somatic recombination events, contributing to chromosomal stability or movement, contributing to allelic exclusion or X chromosome inactivation, or the like.

The methods of the present invention are particularly useful for gene discovery, that is, for identifying, from genomic sequence, regions that function to encode genes, and in a particularly useful embodiment, for identifying regions that function to encode genes not hitherto identified by expression-based or directed cloning and sequencing. In conjunction with verification using the novel single exon microarrays of the present invention, as further described below, the methods herein described become powerful gene discovery tools.

Accordingly, in a preferred embodiment of the present invention, process 25 is used to identify putative coding regions. Two preferred approaches in process 25 for identifying sequence that encodes putative genes are gene prediction and comparative sequence analysis.

Gene prediction can be performed using any of a number of algorithmic methods, embodied in one or more software programs, that identify open reading frames (ORFs) using a variety of heuristics, such as GRAIL, DICTION, and GENEFINDER. Comparative sequence analysis similarly can be performed using any of a variety of known programs that identify regions with lower sequence variability.

As further described in Example 1, below, gene finding software programs yield a range of results. For the newly accessioned human genomic sequence input in Example 1, for example, GRAIL identified the greatest percentage of genomic sequence as putative coding region, 2% of the data analyzed; GENEFINDER was second, calling 1%; and DICTION yielded the least putative coding region, with

0.8% of genomic sequence called as coding region.

Increased reliability can be obtained when consensus is required among several such methods. Although discussed herein particularly with respect to exon calling, consensus among methods will in general increase reliability of predicting other functions as well.

Thus, as indicated by query 26, sequence processing 25, optionally with preprocessing 24, can be repeated with a different method, with consensus among such iterations determined and reported in process 27.

Process 27 compares the several outputs for a given input genomic sequence and identifies consensus among the separately reported results. The consensus itself, as well as the sequence meeting that consensus, is then stored in process 29a, displayed in process 29b, and/or output to process 300 for subsequent identification of a subset thereof suitable for assay.

Multiple levels of consensus can be calculated and reported by process 27. For example, as further described in Example 1, *infra*, process 27 can report consensus as between all specific pairs of methods of gene prediction, as consensus among any one or more of the pairs of methods of gene prediction, or as among all of the gene prediction algorithms used. Thus, in Example 1, process 27 reported that GRAIL and GENEFINDER programs agreed on 0.7% of genomic sequence, that GRAIL and DICTION agreed on 0.5% of genomic sequence, and that the three programs together agreed on 0.25% of the data analyzed. Put another way, 0.25% of the genomic sequence was identified by all three of the programs as containing putative coding region.

Furthermore, consensus can be required among different approaches to identifying a chosen function.

For example, if the function desired to be identified is coding of protein sequence, and a first used approach to exon calling is gene prediction, the process

can be repeated on the same input sequence, or subset thereof, with another approach, such as comparative sequence analysis. In such a case, where comparative sequence analysis follows gene prediction, the comparison can be performed not only on genomic nucleic acid sequence, but additionally or alternatively can be performed on the predicted amino acid sequence translated from the ORFs prior identified by the gene prediction approach.

Although shown as an iterative process, the multiple analyses required to achieve consensus can be done in series, in parallel, or some combination thereof.

Predicted functional sequence, optionally representing a consensus among a plurality of methods and approaches for determination thereof, is passed to process 300 for identification of a subset thereof for functional assay.

In the preferred embodiment of the methods of the present invention, wherein the function sought to be identified is protein coding, process 300 is used to identify a subset thereof suitable for experimental verification by physical and/or bioinformatic approaches.

For example, putative ORFs identified in process 200 can be classified, or binned, bioinformatically into putative genes. This binning can be based *inter alia* upon consideration of the average number of exons/gene in the species chosen for analysis, upon density of exons that have been called on the genomic sequence, and other empirical rules. Thereafter, one or more among the gene-specific ORFs can be chosen for subsequent use in gene expression assay.

Where such subsequent gene expression assay uses amplified nucleic acid, considerations such as desired amplicon length, primer synthesis requirements, putative exon length, sequence GC content, existence of possible secondary structure, and the like can be used to identify

and select those ORFs that appear most likely successfully to amplify. Where subsequent gene expression assay relies upon nucleic acid hybridization, whether or not using amplified product, further considerations involving hybridization stringency can be applied to identify that subset of sequences that will most readily permit sequence-specific discrimination at a chosen hybridization and wash stringency. One particular such consideration is avoidance of putative exons that span repetitive sequence; such sequence can hybridize spuriously to nonspecific message, reducing specific signal in the hybridization.

For bioinformatic assay, there are fewer constraints on the sequences that can be tested experimentally, and in this latter case therefore process 300 can output the entirety of the input sequence.

The subset of sequences identified by process 300 as suitable for use in assay is then used in process 400 to create the physical and/or informational substrate for experimental verification of the predictions made in process 200, and thereafter to assay those substrates.

As mentioned, the methods of the present invention are particularly useful for identifying potential coding regions within genomic sequence. In a preferred embodiment of process 400, therefore, the expression of the sequences predicted to encode protein is verified. The combination of the predictive and experimental methods provides a powerful gene discovery engine.

Thus, in another aspect, the present invention provides methods and apparatus for verifying the expression of putative genes identified within genomic sequence. In particular, the invention provides a novel method of verifying gene expression in which expression of predicted ORFs is measured and confirmed using a novel type of nucleic acid microarray, the genome-derived single exon nucleic acid microarrays of the present invention.

Putative ORFs as predicted by a consensus of gene calling, particularly gene prediction, algorithms in process 200, and as further identified as suitable by process 300, are amplified from genomic DNA using the polymerase chain reaction (PCR). Although PCR is conveniently used, other amplification approaches can also be used.

Amplification schemes can be designed to capture the entirety of each predicted ORF in an amplicon with minimal additional (that is, intronic or intergenic) sequence. Because ORFs predicted from human genomic sequence using the methods of the present invention differ in length, such an approach results in amplicons of varying length.

However, most predicted ORFs are shorter than 500 bp in length, and although amplicons of at least about 100 or 200 base pairs can be immobilized as probes on nucleic acid microarrays, early experimental results using the methods of the present invention have suggested that longer amplicons, at least about 400 or 500 base pairs, are more effective. Furthermore, certain advantages derive from application to the microarray of amplicons of defined size.

Therefore, amplification schemes can alternatively, and preferably, be designed to amplify regions of defined size, preferably at least about 300, 400 or 500 bp, centered about each predicted ORF. Such an approach results in a population of amplicons of limited size diversity, but that typically contain intronic and/or intergenic nucleic acid in addition to putative ORF.

Conversely, somewhat fewer than 10% of ORFs predicted from human genomic sequence according to the methods of the present invention exceed 500 bp in length. Portions of such extended ORFs, preferably at least about 300,400 or 500 bp in length, can be amplified. However, it has been discovered that the percentage success at

amplifying pieces of such ORFs is low, and that such putative exons are more effectively amplified when larger fragments, at least about 1000 or 1500 bp, and even as large as 2000 bp are amplified.

The putative ORFs selected in process 300 are thus input into one or more primer design programs, such as PRIMER3 (available online for use at http://www-genome.wi.mit.edu/cgi-bin/primer/), with a goal of amplifying at least about 500 base pairs of genomic sequence centered within or about ORFs predicted to be no more than about 500 bp, or at least about 1000 - 1500 bp of genomic sequence for ORFs predicted to exceed 500 bp in length, and the primers synthesized by standard techniques. Primers with the requisite sequences can be purchased commercially or synthesized by standard techniques.

Conveniently, a first predetermined sequence can be added commonly to the ORF-specific 5' primer and a second, typically different, predetermined sequence commonly added to each 3' ORF-unique primer. This serves to immortalize the amplicon, that is, serves to permit further amplification of any amplicon using a single set of primers complementary respectively to the common 5' and common 3' sequence elements. The presence of these "universal" priming sequences further facilitates later sequence verification, providing a sequence common to all amplicons at which to prime sequencing reactions. The common 5' and 3' sequences further serve to add a cloning site should any of the ORFs warrant further study.

Such predetermined sequence is usefully at least about 10, 12 or 15 nt in length, and usually does not exceed about 25 nt in length. The "universal" priming sequences used in the examples presented *infra* were each 16 nt long.

The genomic DNA to be used as substrate for amplification will come from the eukaryotic species from

which the genomic sequence data had originally been obtained, or a closely related species, and can conveniently be prepared by well known techniques from somatic or germline tissue or cultured cells of the organism. See, *e.g.*, Short Protocols in Molecular Biology : A Compendium of Methods from Current Protocols in Molecular Biology, Ausubel et al. (eds.), 4^{th} edition (April 1999), John Wiley & Sons (ISBN: 047132938X) and Maniatis *et al.*, Molecular Cloning : A Laboratory Manual, 2^{nd} edition (December 1989), Cold Spring Harbor Laboratory Press (ISBN: 0879693096). Many such prepared genomic DNAs are available commercially, with the human genomic DNAs additionally having certification of donor informed consent.

Although the intronic and intergenic material flanking putative coding regions in the amplicons could potentially interfere with hybridizations during microarray experiments, we have found, surprisingly, that differential expression ratios are not significantly affected. Rather, the predominant effect of exon size is to alter the absolute signal intensity, rather than its ratio. Equally surprising, the art had suggested that single exon probes would not provide sufficient signal intensity for high stringency hybridization analyses; we find that such probes not only provide adequate signal, but have substantial advantages, as herein described.

After partial purification, as by size exclusion spin column, with or without confirmation as to amplicon quality as by gel electrophoresis, each amplicon (single exon probe) is disposed in an array upon a support substrate.

Methods for creating microarrays by deposition and fixation of nucleic acids onto support substrates are well known in the art (Reviewed by Schena et al., see above).

Typically, the support substrate will be glass, although other materials, such as amorphous or crystalline silicon or plastics. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof, can also be used. Typically, the support will be rectangular, although other shapes, particularly circular disks and even spheres, present certain advantages. Particularly advantageous alternatives to glass slides as support substrates for array of nucleic acids are optical discs, as described in WO 98/12559.

The amplified nucleic acids can be attached covalently to a surface of the support substrate or, more typically, applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof.

Robotic spotting devices useful for arraying nucleic acids on support substrates can be constructed using public domain specifications (The MGuide, version 2.0, http://cmgm.stanford.edu/pbrown/mguide/index.html), or can conveniently be purchased from commercial sources (MicroArray GenII Spotter and MicroArray GenIII Spotter, Molecular Dynamics, Inc., Sunnyvale, CA). Spotting can also be effected by printing methods, including those using ink jet technology.

As is well known in the art, microarrays typically also contain immobilized control nucleic acids. For controls useful in providing measurements of background signal for the genome-derived single exon microarrays of the present invention, a plurality of *E. coli* genes can readily be used. As further described in Example 1, 16 or

32 *E. coli* genes suffice to provide a robust measure of background noise in such microarrays.

As is well known in the art, the amplified product disposed in arrays on a support substrate to create a nucleic acid microarray can consist entirely of natural nucleotides linked by phosphodiester bonds, or alternatively can include either nonnative nucleotides, alternative internucleotide linkages, or both, so long as complementary binding can be obtained in the hybridization. If enzymatic amplification is used to produce the immobilized probes, the amplifying enzyme will impose certain further constraints upon the types of nucleic acid analogs that can be generated.

Although particularly described herein as using high density microarrays constructed on planar substrates, the methods of the present invention for confirming the expression of ORFs predicted from genomic sequence can use any of the known types of microarrays, as herein defined, including lower density planar arrays, and microarrays on nonplanar, nonunitary, distributed substrates.

For example, gene expression can be confirmed using hybridization to lower density arrays, such as those constructed on membranes, such as nitrocellulose, nylon, and positively-charged derivatized nylon membranes. Further, gene expression can also be confirmed using nonplanar, bead-based microarrays such as are described in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):166501670 (2000); U.S. Patent No. 6,057,107; and U.S. Patent No. 5,736,330. In theory, a packed collection of such beads provides in aggregate a higher density of nucleic acid probe than can be achieved with spotting or lithography techniques on a single planar substrate.

Planar microarrays on solid substrates, however, provide certain useful advantages, including high throughput and compatibility with existing readers. For

example, each standard microscope slide can include at least 1000, typically at least 2000, preferably 5000 and upto 10,000 - 50,000 or more nucleic acid probes of discrete sequence. The number of sequences deposited will depend on their required application.

Each putative gene can be represented in the array by a single predicted ORF. Alternatively, genes can be represented by more than one predicted ORF. For purposes of measuring differential splicing, more than one predicted ORF will be provided for a putative gene. And as is well known in the art, each probe of defined sequence, representing a single predicted ORF, can be deposited in a plurality of locations on a single microarray to provide redundancy of signal.

The genome-derived single exon microarrays described above differ in several fundamental and advantageous ways from microarrays presently used in the gene expression art, including (1) those created by deposition of mRNA-derived nucleic acids, (2) those created by *in situ* synthesis of oligonucleotide probes, and (3) those constructed from yeast genomic DNA.

Most nucleic acid microarrays that are in use for study of eukaryotic gene expression have as immobilized probes nucleic acids that are derived – either directly or indirectly – from expressed message. As discussed above, it is common, for example, for such microarrays to be derived from cDNA/EST libraries, either from those previously described in the literature, see Lennon *et al.*, or from the *de novo* construction of "problem specific" libraries targeted at a particular biological question, R.S. Thomas *et al.*, *Cancer Res.* (in press). Such microarrays are herein collectively denominated "EST microarrays".

Such EST microarrays by definition can measure expression only of those genes found in EST libraries,

shown herein to represent only a fraction of expressed genes. Furthermore, such libraries — and thus microarrays based thereupon — are biased by the tissue or cell type of message origin, by the expression levels of the respective genes within the tissues, and by the ability of the message successfully to have been reverse-transcribed and cloned.

Thus, as further discussed in Example 1, the methods of the present invention enable sequences that do not appear in EST or other expression databases to be determined - subsequently arrayed for expression measurements could not, therefore, have been represented as probes on an EST microarray. And as further demonstrated in the examples, *infra*, the remaining population of genes identified from genomic sequence by the methods of the present invention — that is, the one third of sequences that had previously been accessioned in EST or other expression databases — are biased toward genes with higher expression levels.

Representation of a message in an EST and/or cDNA library depends upon the successful reverse transcription, optionally but typically with subsequent successful cloning, of the message. This introduces substantial bias into the population of probes available for arraying in EST microarrays..

In contrast, neither reverse transcription nor cloning is required to produce the probes arrayed on the genome-derived single exon microarrays of the present invention. And although the ultimate deposition of a probe on the genome-derived single exon microarray of the present invention depends upon a successful amplification from genomic material, *a priori* knowledge of the sequence of the desired amplicon affords greater opportunity to recover any given probe sequence recalcitrant to amplification than is afforded by the requirement for successful reverse transcription and cloning of unknown message in EST

approaches.

Thus, the genome-derived single exon microarrays of the present invention present a far greater diversity of probes for measuring gene expression, with far less bias, than do EST microarrays presently used in the art.

As a further consequence of their ultimate origin from expressed message, the probes in EST microarrays often contain poly-A (or complementary poly-T) stretches derived from the poly-A tail of mature mRNA. These homopolymeric stretches contribute to cross-hybridization, that is, to a spurious signal occasioned by hybridization to the homopolymeric tail of a labeled cDNA that lacks sequence homology to the gene-specific portion of the probe.

In contrast, the probes arrayed in the genome-derived single exon microarrays of the present invention lack homopolymeric stretches derived from message polyadenylation, and thus can provide more specific signal. Typically, at least about 50, 60 or 75% of the probes on the genome-derived single exon microarrays of the present invention lack homopolymeric regions consisting of A or T, where a homopolymeric region is defined for purposes herein as stretches of 25 or more, typically 30 or more, identical nucleotides.

A further distinction, which also affects the specificity of hybridization, is occasioned by the typical derivation of EST microarray probes from cloned material. Because much of the probe material disposed as probes on EST microarrays is excised or amplified from plasmid, phage, or phagemid vectors, EST microarrays typically include a fair amount of vector sequence, more so when the probes are amplified, rather than excised, from the vector.

In contrast, the vast majority of probes in the genome-derived single exon microarrays of the present invention contain no prokaryotic or bacteriophage vector sequence, having been amplified directly or indirectly from

genomic DNA. Typically, therefore, at least about 50, 60, 70 or 80% or more of individual exon-including probes disposed on a genome-derived single exon microarray of the present invention lack vector sequence, and particularly lack sequences drawn from plasmids and bacteriophage. Preferably, at least about 85, 90 or more than 90% of exon-including probes in the genome-derived single exon microarray of the present invention lack vector sequence. With attention to removal of vector sequences through preprocessing 24, percentages of vector-free exon-including probes can be as high as 95 - 99%. The substantial absence of vector sequence from the genome-derived single exon microarrays of the present invention results in greater specificity during hybridization, since spurious cross-hybridization to a probe vector sequence is reduced.

As a further consequence of excision or amplification of probes from vectors in construction of EST microarrays, the probes arrayed thereon often contain artificial sequence, derived from vector polylinker multiple cloning sites, at both 5' and 3' ends. The probes disposed upon the genome-derived single exon microarrays need have no such artificial sequence appended thereto.

As mentioned above, however, the ORF-specific primers used to amplify putative ORFs can include artificial sequences, typically 5' to the ORF-specific primer sequence, useful for "universal" (that is, independent of ORF sequence) priming of subsequent amplification or sequencing reactions. When such "universal" 5' and/or 3' priming sequences are appended to the amplification primers, the probes disposed upon the genome-derived single exon microarray will include artificial sequence similar to that found in EST microarrays. However, the genome-derived single exon microarray of the present invention can be made without such sequences, and if so constructed, presents an even

smaller amount of nonspecific sequence that would contribute to nonspecific hybridization.

Yet another consequence of typical use of cloned material as probes in EST microarrays is that such microarrays contain probes that result from cloning artifacts, such as chimeric molecules containing coding region of two separate genes. Derived from genomic material, typically not thereafter cloned, the probes of the genome-derived single exon microarrays of the present invention lack such cloning artifacts, and thus provide greater specificity of signal in gene expression measurements.

A further consequence of the cloned origin of probes on many EST microarrays is that the individual probes often have disparate sizes, which can cause the optimal hybridization stringency to vary among probes on a single microarray. In contrast, as discussed above, the probes arrayed on the genome-derived single exon microarrays of the present invention can readily be designed to have a narrow distribution in sizes, with the range of probe sizes no greater than about 10% of the average size, typically no greater than about 5% of the average probe size.

Because of their origin from fully- or partially-spliced message, probes disposed upon EST arrays will often include multiple exons. The percentage of such exon-spanning probes in an EST microarray can be calculated, on average, based upon the predicted number of exons/gene for the given species and the average length of the immobilized probes. For human genes, the near-complete sequence of human chromosome 22, Dunham *et al.*, *Nature* 402(6761):489-95 (1999), predicts that human genes average 5.5 exons/gene. Even with probes of 200 - 500 bp, the vast majority of human EST microarray probes include more than one exon.

In contrast, by virtue of their origin from

algorithmically identified ORFs in genomic sequence, the probes in the genome-derived single exon microarrays of the present invention can consist of individual exons. Thus, in contrast to EST microarrays, at least about 50, 60, 70, 75, 80, 85, 95 or 99% of probes deposited in the genome-derived microarray of the present invention consist of, or include, no more than one predicted ORF.

This provides the ability, not readily achieved using EST microarrays, to use the genome-derived single exon microarrays of the present invention to measure tissue-specific expression of individual exons, which in turn allows differential splicing events to be detected and characterized, and in particular, allows the correlation of differential splicing to tissue-specific expression patterns.

Furthermore, the exons that are represented in EST microarrays are often biased toward the 3' or 5' end of their respective genes, since sequencing strategies used for EST identification are so biased. In contrast, no such 3' or 5' bias necessarily inheres in the selection of exons for disposition on the genome-derived single exon microarrays of the present invention.

Conversely, the probes provided on the genome-derived single exon microarrays of the present invention typically, but need not necessarily, include intronic and/or intergenic sequence that is absent from EST microarrays, which are derived from mature mRNA. Typically, at least about 50, 60, 70, 80 or 90% of the exon-including probes on the genome-derived single exon microarrays of the present invention include sequence drawn from noncoding regions. As discussed above, the additional presence of noncoding region does not significantly interfere with measurement of gene expression, and provides the additional opportunity to assay prespliced RNA, and thus measure such phenomena such as nuclear export control.

The genome-derived single exon microarrays of the present invention are also quite different from *in situ* synthesis microarrays, where probe size is severely constrained by inadequacies in the photolithographic synthesis process.

Typically, probes arrayed on *in situ* synthesis microarrays are limited to a maximum of about 25 bp. As a well known consequence, hybridization to such chips must be performed at low stringency. In order, therefore, to achieve unambiguous sequence-specific hybridization results, the *in situ* synthesis microarray requires substantial redundancy, with concomitant programmed arraying for each probe of probe analogues with altered (*i.e.*, mismatched) sequence.

In contrast, the longer probe length of the genome-derived single exon microarrays of the present invention allows much higher stringency hybridization and wash. Typically, therefore, exon-including probes on the genome-derived single exon microarrays of the present invention average at least about 100, 200, 300, 400 or 500 bp in length. By obviating the need for substantial probe redundancy, this approach permits a higher density of probes for discrete exons or genes to be arrayed on the microarrays of the present invention than can be achieved for *in situ* synthesis microarrays.

A further distinction is that the probes in *in situ* synthesis microarrays typically are covalently linked to the substrate surface. In contrast, the probes disposed on the genome-derived microarray of the present invention typically are, but need not necessarily be, bound noncovalently to the substrate.

Furthermore, the short probe size on *in situ* microarrays causes large percentage differences in the melting temperature of probes hybridized to their complementary target sequence, and thus causes large

percentage differences in the theoretically optimum stringency across the array as a whole.

In contrast, the larger probe size in the microarrays of the present invention create lower percentage differences in melting temperature across the range of arrayed probes.

A further significant advantage of the microarrays of the present invention over *in situ* synthesized arrays is that the quality of each individual probe can be confirmed before deposition. In contrast, the quality of probes cannot be assessed on a probe-by-probe basis for the *in situ* synthesized microarrays presently being used.

The genome-derived single exon microarrays of the present invention are also distinguished over, and present substantial benefits over, the genome-derived microarrays from lower eukaryotes such as yeast. Lashkari *et al.*, *Proc. Natl. Acad. Sci. USA* 94:13057-13062 (1997).

Only about 220 - 250 of the 6100 or so nuclear genes in *Saccharomyces cerevisiae* – that is, only about 4 - 5% – have standard, spliceosomal, introns, Lopez *et al.*, *Nucl. Acids Res.* 28:85-86 (2000); Spingola *et al.*, *RNA* 5(2):221-34 (1999). Furthermore, the entire yeast genome has already been sequenced. These two facts permit the ready amplification and disposition of single-ORF amplicons on such microarray without the requirement for antecedent use of gene prediction and/or comparative sequence analyses.

Thus, a significant aspect of the present invention is the ability to identify and to confirm expression of predicted coding regions in genomic sequence drawn from eukaryotic organisms that have a higher percentage of genes having introns than do yeast such as *Saccharomyces cerevisiae*, particularly in genomic sequence drawn from eukaryotes in which at least about 10, 20 or 50%

of protein-encoding genes have introns. In preferred embodiments, the methods and apparatus of the present invention are used to identify and confirm expression of novel genes from genomic sequence of eukaryotes in which the average number of introns per gene is at least about one, two or three or more.

After the physical substrate is prepared, experimental verification of predicted function is performed.

In a preferred embodiment of the present invention, where the function sought to be identified in genomic sequence is protein coding, experimental verification is performed by measuring expression of the putative ORFs, typically through nucleic acid hybridization experiments, and in particularly preferred embodiments, through hybridization to genome-derived single exon microarrays prepared as above- described.

Expression is conveniently measured and expressed for each probe in the microarray as a ratio of the expression measured concurrently in a plurality of mRNA sources, according to techniques well known in the microarray art, *Reviewed in* Schena et al., and as further described in Example 2, below. The mRNA source for the reference against which specific expression is measured can be drawn from a homogeneous mRNA source, such as a single cultured cell-type, or alternatively can be heterogeneous, as from a pool of mRNA derived from multiple tissues and/or cell types, as further described in Example 2, *infra*.

mRNA can be prepared by standard techniques, see Ausubel et al. and Maniatis *et al.*, or purchased commercially. The mRNA is then typically reverse-transcribed in the presence of labeled nucleotides: the index source (that in which expression is desired to be measured) is reverse transcribed in the presence of nucleotides labeled with a first label, typically a

fluorophore (fluorochrome; fluor; fluorescent dye); the reference source is reverse transcribed in the presence of a second label, typically a fluorophore, typically fluorometrically-distinguishable from the first label. As further described in Example 2, *infra*, Cy3 and Cy5 dyes prove particularly useful in these methods. After partial purification of the index and reference targets, hybridization to the probe array is conducted according to standard techniques, typically under a coverslip.

After wash, microarrays are conveniently scanned using a commercial microarray scanning device, such as a Gen3 Scanner (Molecular Dynamics, Sunnyvale, CA). Data on expression is then passed, with or without interim storage, to process 500, where the results for each probe are related to the original sequence.

Often, hybridization of target material to the genome-derived single exon microarray will identify certain of the probes thereon as of particular interest. Thus, it is often desirable that the user be able readily to obtain sufficient quantities of an individual probe, either for subsequent arrayed deposition upon an additional support substrate, often as part of a microarray having a plurality of probes so identified, or alternatively or additionally as a solitary solid-phase or solution-phase probe, for further use.

Thus, in another aspect, the present invention provides compositions and kits for the ready production of nucleic acids identical in sequence to, or substantially identical in sequence to, probes on the genome-derived single exon microarrays of the present invention.

In this aspect, a small quantity of each probe is disposed, typically without attachment to substrate, in a spatially-addressable ordered set, typically one per well of a microtiter dish. Although a 96 well microtiter plate can be used, greater efficiency is obtained using higher

density arrays, such as are provided by microtiter plates having 384, 864, 1536, 3456, 6144, or 9600 wells, and although microtiter plates having physical depressions (wells) are conveniently used, any device that permits addressable withdrawal of reagent from fluidly-noncommunicating areas can be used.

In this aspect of the invention, therefore, a fluidly noncommunicating addressable ordered set of individual probes, corresponding to those on a genome-derived single exon microarray, is provided, with each probe in sufficient quantity to permit amplification, such as by PCR. As earlier mentioned, the ORF-specific 5' primers used for genomic amplification can have a first common sequence added thereto, and the ORF-specific 3' primers used for genomic amplification can have a second, different, common sequence added thereto, thus permitting, in this preferred embodiment, the use of a single set of 5' and 3' primers to amplify any one of the probes from the amplifiable ordered set.

Each discrete amplifiable probe can also be packaged with amplification primers, solutes, buffers, *etc.*, and can be provided in dry (*e.g.*, lyophilized) form or wet, in the latter case typically with addition of agents that retard evaporation.

In another aspect of the present invention, a genome-derived single-exon microarray is packaged together with such an ordered set of amplifiable probes corresponding to the probes, or one or more subsets of probes, thereon. In alternative embodiments, the ordered set of amplifiable probes is packaged separately from the genome-derived single exon microarray.

In some embodiments, the microarray and/or ordered probe set are further packaged with recordable media that provide probe identification and addressing information, and that can additionally contain annotation

information, such as gene expression data. Such recordable media can be packaged with the microarray, with the ordered probe set, or with both.

If the microarray is constructed on a substrate that incorporates recordable media, such as is described in international patent application no. WO 98/12559, then separate packaging of the genome-derived single exon microarray and the bioinformatic information is not required.

The amount of amplifiable probe material should be sufficient to permit at least one amplification sufficient for subsequent hybridization assay.

Although the use of high density genome-derived microarrays on solid planar substrates is presently a preferred approach for the physical confirmation and characterization of the expression of sequences predicted to encode protein, other types of microarrays (as herein defined) can also be used.

Furthermore, as earlier mentioned, experimental verification of the function predicted from genomic sequence in process 200 can be bioinformatic, rather than, or additional to, physical verification.

For example, where the function desired to be identified is protein coding, the predicted ORFs can be compared bioinformatically to sequences known or suspected of being expressed.

Thus, the sequences output from process 300 (or process 200), can be used to query expression databases, such as EST databases, SNP ("single nucleotide polymorphism") databases, known cDNA and mRNA sequences, SAGE ("serial analysis of gene expression") databases, and more generalized sequence databases that allow query for expressed sequences. Such query can be done by any sequence query algorithm, such as BLAST ("basic local alignment search tool"). The results of such query –

including information on identical sequences and information on nonidentical sequences that have diffuse or focal regions of sequence homology to the query sequence – can then be passed directly to process 500, or used to inform analyses subsequently undertaken in process 200, process 300, or process 400.

Experimental data, whether obtained by physical or bioinformatic assay in process 400, is passed to process 500 where it is usefully related to the sequence data itself, a process colloquially termed "annotation". Such annotation can be done using any technique that usefully relates the functional information to the sequence, as, for example, by incorporating the functional data into the record itself, by linking records in a hierarchical or relational database, by linking to external databases, or by a combination thereof. Such database techniques are well within the skill in the art.

The annotated sequence data can be stored locally, uploaded to genomic sequence database 100, and/or displayed 800.

The methods and apparatus of the present invention rapidly produce functional information from genomic sequence. Coupled with the escalating pace at which sequence now accumulates, the rapid pace of sequence annotation produces a need for methods of displaying the information in meaningful ways.

FIG. 3 shows visual display 80 presenting a single genomic sequence annotated according to the present invention. Because of its nominal resemblance to artistic works of Piet Mondrian, visual display 80 is alternatively described herein as a "Mondrian".

Each of the visual elements of display 80 is aligned with respect to the genomic sequence being annotated (hereinafter, the "annotated sequence"). Given the number of nucleotides typically represented in an

annotated sequence, representation of individual nucleotides would rarely be readable in hard copy output of display 80. Typically, therefore, the annotated sequence is schematized as rectangle 89, extending from the left border of display 80 to its right border. By convention herein, the left border of rectangle 89 represents the first nucleotide of the sequence and the right border of rectangle 89 represents the last nucleotide of the sequence.

As further discussed below, however, the Mondrian visual display of annotated sequence can serve as a convenient graphical user interface for computerized representation, analysis, and query of information stored electronically. For such use, the individual nucleotides can conveniently be linked to the X axis coordinate of rectangle 89. This permits the annotated sequence at any point within rectangle 89 readily to be viewed, either automatically – for example, by time-delayed appearance of a small overlaid window upon movement of a cursor or other pointer over rectangle 89 – or through user intervention, as by clicking a mouse or other pointing device at a point in rectangle 89.

Visual display 80 is generated after user specification of the genomic sequence to be displayed. Such specification can consist of or include an accession number for a single clone (*e.g.*, a single BAC accessioned into GenBank), wherein the starting and stopping nucleotides are thus absolutely identified, or alternatively can consist of or include an anchor or fulcrum point about which a chosen range of sequence is anchored, thus providing relative endpoints for the sequence to be displayed. For example, the user can anchor such a range about a given chromosomal map location, gene name, or even a sequence returned by query for similarity or identity to an input query sequence. When visual

display 80 is used as a graphical user interface to computerized data, additional control over the first and last displayed nucleotide will typically be dynamically selectable, as by use of standard zooming and/or selection tools.

Field 81 of visual display 80 is used to present the output from process 200, that is, to present the bioinformatic prediction of those sequences having the desired function within the genomic sequence. Functional sequences are typically indicated by at least one rectangle 83 (83a, 83b, 83c), the left and right borders of which respectively indicate, by their X-axis coordinates, the starting and ending nucleotides of the region predicted to have function.

Where a single bioinformatic method or approach identifies a plurality of regions having the desired function, a plurality of rectangles 83 is disposed horizontally in field 81. Where multiple methods and/or approaches are used to identify function, each such method and/or approach can be represented by its own series of horizontally disposed rectangles 83, each such horizontally disposed series of rectangles offset vertically from those representing the results of the other methods and approaches.

Thus, rectangles 83a in FIG. 3 represent the functional predictions of a first method of a first approach for predicting function, rectangles 83b represent the functional predictions of a second method and/or second approach for predicting that function, and rectangles 83c represent the predictions of a third method and/or approach.

Where the function desired to be identified is protein coding, field 81 is used to present the bioinformatic prediction of sequences encoding protein. For example, rectangles 83a can represent the results from

GRAIL or GRAIL II, rectangles 83b can represent the results from GENEFINDER, and rectangles 83c can represent the results from DICTION.

Optionally, and preferably, rectangles 83 collectively representing predictions of a single method and/or approach are identically colored and/or textured, and are distinguishable from the color and/or texture used for a different method and/or approach.

Alternatively, or in addition, the color, hue, density, or texture of rectangles 83 can be used further to report a measure of the bioinformatic reliability of the prediction. For example, many gene prediction programs will report a measure of the reliability of prediction. Thus, increasing degrees of such reliability can be indicated, *e.g.*, by increasing density of shading. Where display 80 is used as a graphical user interface, such measures of reliability, and indeed all other results output by the program, can additionally or alternatively be made accessible through linkage from individual rectangles 83, as by time-delayed window ("tool tip" window), or by pointer (*e.g.*, mouse)-activated link.

As earlier described, increased predictive reliability can be achieved by requiring consensus among methods and/or approaches to determining function. Thus, field 81 can include a horizontal series of rectangles 83 that indicate one or more degrees of consensus in predictions of function.

Although FIG. 3 shows three series of horizontally disposed rectangles in field 81, display 80 can include as few as one such series of rectangles and as many as can discriminably be displayed, depending upon the number of methods and/or approaches used to predict a given function.

Furthermore, field 81 can be used to show predictions of a plurality of different functions.

However, the increased visual complexity occasioned by such display makes more useful the ability of the user to select a single function for display. When display 80 is used as a graphical user interface for computer query and analysis, such function can usefully be indicated and user-selectable, as by a series of graphical buttons or tabs (not shown in FIG. 3).

Rectangle 89 is shown in FIG. 3 as including interposed rectangle 84. Rectangle 84 represents the portion of annotated sequence for which predicted functional information has been assayed physically, with the starting and ending nucleotides of the assayed material indicated by the X axis coordinates of the left and right borders of rectangle 84. Rectangle 85, with optional inclusive circles 86 (86a, 86b, and 86c) displays the results of such physical assay.

Although a single rectangle 84 is shown in FIG. 3, physical assay is not limited to just one region of annotated genomic sequence. It is expected that an increasing percentage of regions predicted to have function by process 200 will be assayed physically, and that display 80 will accordingly, for any given genomic sequence, have an increasing number of rectangles 84 and 85, representing an increased density of sequence annotation.

Where the function desired to be identified is protein coding, rectangle 84 identifies the sequence of the probe used to measure expression. In embodiments of the present invention where expression is measured using genome-derived single exon microarrays, rectangle 84 identifies the sequence included within the probe immobilized on the support surface of the microarray. As noted *supra*, such probe will often include a small amount of additional, synthetic, material incorporated during amplification and designed to permit reamplification of the probe, which sequence is typically not shown in display 80.

Rectangle 87 is used to present the results of bioinformatic assay of the genomic sequence. For example, where the function desired to be identified is protein coding, process 400 can include bioinformatic query of expression databases with the sequences predicted in process 200 to encode exons. And as earlier discussed, because bioinformatic assay presents fewer constraints than does physical assay, often the entire output of process 200 can be used for such assay, without further subsetting thereof by process 300. Therefore, rectangle 87 typically need not have separate indicators therein of regions submitted for bioinformatic assay; that is, rectangle 87 typically need not have regions therein analogous to rectangles 84 within rectangle 89.

Rectangle 87 as shown in FIG. 3 includes smaller rectangles 880 and 88. Rectangles 880 indicate regions that returned a positive result in the bioinformatic assay, with rectangles 88 representing regions that did not return such positive results. Where the function desired to be predicted and displayed is protein coding, rectangles 880 indicate regions of the predicted exons that identify sequence with significant similarity in expression databases, such as EST, SNP, SAGE databases, with rectangles 88 indicating genes novel over those identified in existing expression data bases.

Rectangles 880 can further indicate, through color, shading, texture, or the like, additional information obtained from bioinformatic assay.

For example, where the function assayed and displayed is protein coding, the degree of shading of rectangles 880 can be used to represent the degree of sequence similarity found upon query of expression databases. The number of levels of discrimination can be as few as two (identity, and similarity, where similarity has a user-selectable lower threshold). Alternatively, as

many different levels of discrimination can be indicated as can visually be discriminated.

Where display 80 is used as a graphical user interface, rectangles 880 can additionally provide links directly to the sequences identified by the query of expression databases, and/or statistical summaries thereof. As with each of the precedingly-discussed uses of display 80 as a graphical user interface, it should be understood that the information accessed via display 80 need not be resident on the computer presenting such display, which often will be serving as a client, with the linked information resident on one or more remotely located servers.

Rectangle 85 displays the results of physical assay of the sequence delimited by its left and right borders.

Rectangle 85 can consist of a single rectangle, thus indicating a single assay, or alternatively, and increasingly typically, will consist of a series of rectangles (85a, 85b, 85c) indicating separate physical assays of the same sequence.

Where the function assayed is gene expression, and where gene expression is assayed as herein described using simultaneous two-color fluorescent detection of hybridization to genome-derived single exon microarrays, individual rectangles 85 can be colored to indicate the degree of expression relative to control. Conveniently, shades of green can be used to depict expression in the sample over control values, and shades of red used to depict expression less than control, corresponding to the spectra of the Cy3 and Cy5 dyes conventionally used for respective labeling thereof. Additional functional information can be provided in the form of circles 86 (86a, 86b, 86c), where the diameter of the circle can be used to indicate expression intensity. As discussed *infra*, such

relative expression (expression ratios) and absolute expression (signal intensity) can be expressed using normalized values.

Where display 80 is used as a graphical user interface, rectangle 85 can be used as a link to further information about the assay. For example, where the assay is one for gene expression, each rectangle 85 can be used to link to information about the source of the hybridized mRNA, the identity of the control, raw or processed data from the microarray scan, or the like.

FIG. 4 is rendition of display 80 representing gene prediction and gene expression for a hypothetical BAC, showing conventions used in the Examples presented *infra*. BAC sequence ("Chip seq.") 89 is presented, with the physically assayed region thereof (corresponding to rectangle 84 in FIG. 3) shown in white. Algorithmic gene predictions are shown in field 81, with predictions by GRAIL shown, predictions by GENEFINDER, and predictions by DICTION shown. Within rectangle 87, regions of sequence that, when used to query expression databases, return identical or similar sequences ("EST hit") are shown as white rectangles (corresponding to rectangles 880 in FIG. 3), gray indicates low homology, and black indicates unknowns (where black and gray would correspond to rectangles 88 in FIG. 3).

Although FIGS. 3 and 4 show a single stretch of sequence, uninterrupted from left to right, longer sequences are usefully represented by vertical stacking of such individual Mondrians, as shown in FIGS. 9 and 10.

Single Exon Probes Useful For Measuring Gene Expression

The methods and apparatus of the present invention rapidly produce functional information from genomic sequence. Where the function to be identified is

protein coding, the methods and apparatus of the present invention rapidly identify and confirm the expression of portions of genomic sequence that function to encode protein. As a direct result, the methods and apparatus of the present invention rapidly yield large numbers of single-exon nucleic acid probes, the majority from previously unknown genes, each of which is useful for measuring and/or surveying expression of a specific gene in one or more tissues or cell types.

It is, therefore, another aspect of the present invention to provide genome-derived single exon nucleic acid probes useful for gene expression analysis, and particularly for gene expression analysis by microarray.

Using the methods and genome-derived single-exon microarrays of the present invention, we have for example readily identified a large number of unique ORFs from human genomic sequence. Using single exon probes that encompass these ORFs, we have demonstrated, through microarray hybridization analysis, the expression of 12,821 of these ORFs in brain.

As would immediately be appreciated by one of skill in the art, each single exon probe having demonstrable expression in brain is currently available for use in measuring the level of its ORF's expression in brain.

Diseases of the brain and nervous system are a significant cause of human morbidity and mortality. Increasingly, genetic factors are being found that contribute to predisposition, onset, and/or aggressiveness of most, if not all, of these diseases. Although mutations in single genes have been identified as causative for some diseases of the brain and nervous system, for the most part these disorders are believed to have polygenic etiologies.

For example, over the past few decades Alzheimer's disease (AD), once considered a rare disorder,

has become recognized as a major public health problem; over 4,000,000 people in the United States are now estimated to suffer with various stages of this progressive, degenerative brain disorder.

Although there is no agreement on the exact incidence or prevalence of Alzheimer's disease, in part due to varying diagnostic criteria and difficulties of differential diagnosis among dementias, the studies are consistent in pointing to an exponential rise in prevalence of this disease with age. After age 65, the percentage of affected people approximately doubles with every decade of life, regardless of definition. Among people age 85 or older, studies suggest that 25 to 35 percent have dementia, including Alzheimer's disease; one study reports that 47.2 percent of people over age 85 have Alzheimer's disease, exclusive of other dementias.

Alzheimer's disease progressively destroys memory, reason, judgment, language, and, eventually, the ability to carry out even the simplest of tasks. Anatomic changes associated with Alzheimer's disease begin in the entorhinal cortex, proceed to the hippocampus, and then gradually spread to other regions, particularly the cerebral cortex. Chief among such anatomic changes are the presence of characteristic extracellular plaques and internal neurofibrillary tangles.

Alzheimer's disease has been suspected to have a multifactorial genetic etiological component for almost half a century. Sjogren et al., Acta Psychiat. Neurol. Scand. 82(suppl.): 1-152 (1952).

At least four genes have been identified to date that contribute to development of Alzheimer's disease: AD1 is caused by mutations in the amyloid precursor gene (APP); AD2 is associated with the APOE4 allele on chromosome 19; AD3 is caused by mutation in a chromosome 14 gene encoding a 7-transmembrane domain protein, presenilin-1 (PSEN1), and

AD4 is caused by mutation in a gene on chromosome 1 that encodes a similar 7-transmembrane domain protein, presenilin-2 (PSEN2).

There is strong evidence, however, for additional, as yet uncharacterized, AD loci on other chromosomes.

For example, Daw et al., Am. J. Hum. Genet. 66: 196-204 (2000), estimated the number of additional quantitative trait loci (QTLs) and their contribution to the variance in age at onset of AD, and reported that 4 loci make a contribution to the variance in age at onset of late-onset AD similar to or greater in magnitude than that made by apoE, with one locus making a contribution several times greater than that of apoE. These results suggest that several genes not yet localized may play a larger role than does apoE in late-onset AD.

In accord, three groups recently announced the possible existence of an AD susceptibility gene on chromosome 10. Bertram et al., Science 290(5500):2302-2303 (2000); Ertekin-Taner et al., Science 290(5500):2303-2304 (2000); and Myers et al., Science 290(5500):2304-23055 (2000).

As another example, multiple sclerosis (MS) affects about 350,000 Americans, with approximately 200 new cases diagnosed each week, with an estimated annual monetary cost in the U.S. alone of $2.5 billion.

Clinically, MS is an unpredictable disorder, with symptoms, presentation and course falling broadly into one of several clinical patterns. In relapsing-remitting (RR) MS, the disease first manifests as a series of attacks followed by complete or partial remissions, with symptoms returning later after a period of stability. In primary-progressive (PP) MS, there is a gradual clinical decline with no distinct remissions, although there may be

temporary plateaus or minor relief from symptoms. Secondary-progressive (SP) MS begins with a relapsing-remitting course followed by a later primary-progressive course. Rarely, patients may have a progressive-relapsing (PR) course in which the disease takes a progressive path punctuated by acute attacks. PP., SP, and PR MS are sometimes lumped together and called chronic progressive MS. The waxing and waning course characteristic of RR, SP and PR MS makes differential diagnosis difficult.

Anatomically, MS attacks are associated with focal inflammation in areas of the white matter of the central nervous system (CNS), accompanied or followed by demyelination in these areas, termed plaques. Destruction of the myelin sheath slows or blocks neurological transmission, leading to diminished or lost function. Clinical manifestations depend upon the location of the plaques and severity of demyelination, and range from fatigue, the most common symptom of MS, to visual impairment, due to inflammation of the optic nerve, termed optic neuritis, to numbness and paresthesias, to focal muscular weakness, ataxia, and bladder incontinence.

Increasing evidence suggests that genotype contributes to susceptibility to MS.

As early as 1965, McAlpine, in Multiple Sclerosis: A Reappraisal (McAlpine, ed.), Williams and Wilkins Co. pp. 61-74 (1965), concluded that the risk to a first-degree relative of a patient with multiple sclerosis is at least 15 times that for a member of the general population, but could discern no definite genetic pattern of inheritance.

Subsequently, many studies associated MS with HLA (MHC) haplotype. Haines et al., Hum. Molec. Genet. 7:1229-1234 (1998), studying a data set of 98 multiplex MS families, confirmed earlier reports that genetic linkage to the MHC can be explained by association with the HLA-DR2

allele, but suggested that MHC association explains only between 17% and 62% of the genetic etiology of MS.

From a review of genomic screens, Dyment et al., Hum. Molec. Genet. 6: 1693-1698 (1997), concluded that a number of genes with interacting effects are likely and that no single region has a major influence on familial risk. Chataway et al., Brain 121: 1869-1887 (1998), reporting a follow-up on U.K. studies using a systematic genome screen to determine the genetic basis of MS, stated that a gene of major effect had been excluded from 95% of the genome and one with a moderate role from 65%, results thus suggesting that multiple sclerosis depends on independent or epistatic effects of several genes, each with small individual effects, rather than a very few genes of major biologic importance.

As a yet further example, schizophrenia has long been recognized to have complex, likely polygenic, genetic contributions.

Schizophrenia is a common psychiatric disorder, occurring in 1 to 1.5 percent of the population worldwide, and is characterized by variable constellations of symptoms drawn from a universe of behavioral abnormalities. Although there are accepted alternative diagnostic criteria, primary criteria for diagnosis require two or more of the following, each present for a significant portion of time during a 1-month period (or less if successfully treated): (1) delusions; (2) hallucinations ; (3) disorganized speech (e.g., frequent derailment or incoherence); (4) grossly disorganized or catatonic behavior; (5) negative symptoms, i.e., affective flattening, alogia, or avolition. (Diagnostic and Statistic Manual of Mental Disorders DSM-IV-TR, American Psychiatric Association (2000)). Only one such symptom is required if delusions are bizarre or hallucinations consist of a voice keeping up a running commentary on the

person's behavior or thoughts, or consist of two or more voices conversing with each other.

Three-quarters of persons with schizophrenia develop the disease between 16 and 25 years of age: onset is uncommon after age 30, rare after age 40. In the 16 to 25 year old age group, schizophrenia affects more men than women; in the 25-30 year old group, the incidence is higher in women than in men. Studies have shown that some persons with schizophrenia recover completely, and many others improve to the point where they can live independently, often with the maintenance of drug therapy. However, approximately 15 percent of people with schizophrenia respond only moderately to medication and require extensive support throughout their lives, while another 15 percent simply do not respond to existing treatment.

Schizophrenia has long been known to have a significant genetic component. Studies have consistently demonstrated that the risk to relatives of a proband with schizophrenia is higher than the risk to relatives of controls. Moldin, in Genetics and Mental Disorders: Report of the NIMH Genetics Workgroup (NIH publication 98-4268, (1998), reviewed family and twin studies published between 1920 and 1987 and found the recurrence risk ratios to be 48 for monozygotic twins, 11 for first-degree relatives, 4.25 for second-degree relatives, and 2 for third-degree relatives. He also found that concordance rates for monozygotic twins averaged 46%, even when reared in different families, whereas the concordance rates for dizygotic twins averaged only 14%. The prevalence of schizophrenia is known to be higher in biologic than in adoptive relatives of schizophrenic adoptees.

The mode of inheritance is unclear, however. Susceptibility has been mapped to many loci, including chromosomes 1q21-q22, 5, 6p23, 8p22-p21, 11q, 13q14-q21, 13q32, 15q15, 15q14, 18p, and 22q11. Chromosome

19 has also been implicated in schizophrenia, at 2 different sites, as have sites on the X chromosome. Wei et al., Nature Genet. 25:376-377 (2000) report more specifically that the NOTCH4 locus is associated with susceptibility to schizophrenia.

In general, however, it is believed that development of schizophrenia involves multiple loci.

For example, Williams et al., Hum. Molec. Genet. 8:1729-1739 (1999) undertook a systematic search for linkage in 196 affected sib pairs (ASPs) with schizophrenia. Using 229 microsatellite markers at an average intermarker distance of 17.26 cM, followed in a second stage by a further 54 markers allowing the regions identified in stage 1 to be typed at an average spacing of 5.15 cM, Williams et al. considered results on chromosomes 4p, 18q, and Xcen as suggestive; however, given the scores, Williams et al. interpreted their results as suggesting that common genes of major effect (susceptibility ratio more than 3) are unlikely to exist for schizophrenia.

Similarly, Shaw et al., Am. J. Med. Genet. 81(5):364-76 (1998), in a genome-wide search for schizophrenia susceptibility genes, found that twelve chromosomes (1, 2, 4, 5, 8, 10, 11, 12, 13, 14, 16, and 22) had at least one region with a nominal P value <0.05, that two of these chromosomes had a nominal P value <0.01 (chromosomes 13 and 16), and that five chromosomes (1, 2, 4, 11, and 13) had at least one marker with a lod score >2.0, suggesting the existence of multiple loci that contribute to schizophrenia susceptibility.

As yet another example, multiple genes are thought to predispose to epilepsy.

Epilepsy is characterized by recurrent, paroxysmal disorders of cerebral function (seizures); that is, by sudden, brief attacks of altered consciousness, motor activity, sensory phenomena, or inappropriate

behavior. The risk of developing epilepsy is 1% in the period from birth to age 20, and 3% at age 75.

Epilepsy is caused by excessive discharge of cerebral neurons. Clinical manifestations depend on the type and location of discharge. In partial seizures, for example, the excess neuronal discharge is contained within one region of the cerebral cortex. Simple partial seizures consist of motor, sensory, or psychomotor phenomena without loss of consciousness; the specific phenomenon reflects the affected area of the brain. In generalized seizures, the discharge bilaterally and diffusely involves the entire cortex. Sometimes a focal lesion of one part of a hemisphere activates the entire cerebrum bilaterally so rapidly that it produces a generalized tonic-clonic seizure before a focal sign appears.

Epilepsy is a family of disorders. Those that are idiopathic are believed to have multiple genetic contributions. For example, idiopathic generalized epilepsy (IGE) is characterized by recurring generalized seizures in the absence of detectable brain lesions and/or metabolic abnormalities. Twin and family studies suggest that genetic factors play a key part in its etiology. Although a mutation in the CACNB4 gene can cause the disorder, linkage to 8q24, Zara et al., Hum. Molec. Genet. 4: 1201-1207(1995), 3q26 and 14q23, Sander et al., Hum. Molec. Genet. 9:1465-1472 (2000), and 2q36 has been also demonstrated, with a multilocus model appearing to fit best the observed familial patterns.

Polygenic contributions to the etiology of various neurologic cancers have similarly been described.

For example, gliomas account for 45% of intracranial tumors, and multiple loci have been implicated in its development, with losses of chromosome 17p, increase in copy number of chromosome 7, structural abnormalities of

chromosomes 9p and 19q, and genes on chromosome 10 among the suspects.

Other significant diseases of brain and nervous tissue are also believed to have a genetic, typically polygenic, etiologic component. These diseases include, for example, Parkinson's disease, dementia with Lewy bodies, frontotemporal dementia, corticobasal ganglionic degeneration, progressive supranuclear palsy, prion diseases (Creutzfeld-Jakob, Gerstmann-Strausller-Shenker, familial fatal insomnia), Tourette's Syndrome, corticobasal degeneration, multiple system atrophy, striatonigral degeneration, Shy-Drager syndrome, olivopontocerebellar atrophy, spinocerebellar ataxia, Friedreich ataxia, ataxia-telangiectasia, amyotrophic lateral sclerosis, bulbospinal atrophy (Kennedy's syndrome), spinal muscular atrophy, neuronal storage diseases (sphingolipid, mucopolysaccharide, mucolipid), leukodystrophy, Krabbe disease, metachromic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, Canavan disease, mitochondrial encephalomyopathy, Leigh disease, neurofibromatosis (Type 1 and Type II), tuberous sclerosis, paraneoplastic syndrome, subacute cerebellar degeneration, subacute sensory neuropathy, opsoclonus/myoclonus, retinal degeneration, stiff-man syndrome and Von Hippel-Lindau disease.

Many neurologic cancers other than gliomas have also been shown or suspected to have genetic bases or contributions. Among these cancers are astrocytoma, fibrillary astrocytoma, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, oligodendroglioma, ependymoma, gangliocytoma, ganglioglioma, medulloblastoma, primary brain germ cell tumor, pineocytoma, pineoblastoma, and meningioma.

Other disorders of brain and central nervous system that likely have genetic components include the

various forms of neural deafness, catatonia, depression, bipolar (manic-depressive) disorder, Wilson's Disease, Pick disease, neuromyelitis optica (Devic disease), central pontine myelinolysis, Marchiafava-Bignami disease, Guillain-Barre syndrome, sleep disorders (insomnia, myoclonus, narcolepsy, cataplexy, sleep apnea), amnesia, aphasias (including Broca's aphasia and Wernicke's aphasia), cortical blindness, visual agnosia, auditory agnosia, and Kluver-Bucy syndrome.

The human genome-derived single exon nucleic acid probes and microarrays of the present invention are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of human brain, particularly those diseases with polygenic etiology. With each of the single exon probes described herein shown to be expressed at detectable levels in human brain, and with about 2/3 of the probes identifying novel genes, the single exon microarrays of the present invention provide exceptionally high informational content for such studies.

For example, diagnosis (including differential diagnosis among clinically indistinguishable disorders), staging, and/or grading of a disease can be based upon the quantitative relatedness of a patient gene expression profile to one or more reference expression profiles known to be characteristic of a given neurologic disease, or to specific grades or stages thereof.

In one embodiment, the patient gene expression profile is generated by hybridizing nucleic acids obtained directly or indirectly from transcripts expressed in the patient's brain (or other CNS tissues, including cultured tissues) to the genome-derived single exon microarray of the present invention. Reference profiles are be obtained similarly by hybridizing nucleic acids from individuals with known disease. Methods for quantitatively relating gene expression profiles, without regard to the function of

the protein encoded by the gene, are disclosed in WO 99/58720, incorporated herein by reference in its entirety.

In another approach, the genome-derived single exon probes and microarrays of the present invention can be used to interrogate genomic DNA, rather than pools of expressed message; this latter approach permits predisposition to and/or prognosis of neurologic disease to be assessed through the massively parallel determination of altered copy number, deletion, or mutation in the patient's genome of exons known to be expressed in human brain. The algorithms set forth in WO 99/58720 can be applied to such genomic profiles without regard to the function of the protein encoded by the interrogated gene.

The utility is specific to the probe; at sufficiently high hybridization stringency, which stringencies are well known in the art – see Ausubel et al. and Maniatis *et al.* – each probe reports the level of expression of message specifically containing that ORF.

It should be appreciated, however, that the probes of the present invention, for which expression in the brain has been demonstrated are useful for both measurement in the brain and for survey of expression in other tissues.

Significant among such advantages is the presence of probes for novel genes.

As mentioned above and further detailed in Examples 1 and 2, the methods described enable ORFs which are not present in existing expression databases to be identified. And the fewer the number of tissues in which the ORF can be shown to be expressed, the more likely the ORF will prove to be part of a novel gene: as further discussed in Example 2, ORFs whose expression was measurable in only a single of the tested tissues were represented in existing expression databases at a rate of only 11%, whereas 36% of ORFs whose expression was

measurable in 9 tissues were present in existing expression databases, and fully 45% of those ORFs expressed in all ten tested tissues were present in existing expressed sequence databases.

Either as tools for measuring gene expression or tools for surveying gene expression, the genome-derived single exon probes of the present invention have significant advantages over the cDNA or EST-based probes that are currently available for achieving these utilities.

The genome-derived single exon probes of the present invention are useful in constructing genome-derived single exon microarrays; the genome-derived single exon microarrays, in turn, are useful devices for measuring and for surveying gene expression in the human.

Gene expression analysis using microarrays – conventionally using microarrays having probes derived from expressed message – is well-established as useful in the biological research arts (see Lockhart et al. Nature 405, 827-836).

Microarrays have been used to determine gene expression profiles in cells in response to drug treatment (see, for example, Kaminski *et al.*, "Global Analysis of Gene Expression in Pulmonary Fibrosis Reveals Distinct Programs Regulating Lung Inflammation and Fibrosis," *Proc. Natl. Acad. Sci. USA* 97(4):1778-83 (2000); Bartosiewicz *et al.*, "Development of a Toxicological Gene Array and Quantitative Assessment of This Technology," *Arch. Biochem. Biophys.* 376(1):66-73 (2000)), viral infection (see for example, Geiss *et al.*, "Large-scale Monitoring of Host Cell Gene Expression During HIV-1 Infection Using cDNA Microarrays," *Virology* 266(1):8-16 (2000)) and during cell processes such as differentiation, senescence and apoptosis (see, for example, Shelton *et al.*,"Microarray Analysis of Replicative Senescence," *Curr. Biol.* 9(17):939-45 (1999); Voehringer *et al.*, "Gene Microarray Identification of Redox

and Mitochondrial Elements That Control Resistance or Sensitivity to Apoptosis," Proc. Natl. Acad. Sci. USA 97(6):2680-5 (2000)).

Microarrays have also been used to determine abnormal gene expression in diseased tissues (see, for example, Alon *et al.*, "Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays," *Proc. Natl. Acad. Sci. USA* 96(12):6745-50 (1999); Perou *et al.*, "Distinctive Gene Expression Patterns in Human Mammary Epithelial Cells and Breast Cancers, *Proc. Natl. Acad. Sci. USA* 96(16):9212-7 (1999); Wang *et al.*, "Identification of Genes Differentially Over-expressed in Lung Squamous Cell Carcinoma Using Combination of cDNA Subtraction and Microarray Analysis," *Oncogene* 19(12):1519-28 (2000); Whitney *et al.*, "Analysis of Gene Expression in Multiple Sclerosis Lesions Using cDNA Microarrays," *Ann. Neurol.* 46(3):425-8 (1999)), in drug discovery screens (see, for example, Scherf *et al.*, "A Gene Expression Database for the Molecular Pharmacology of Cancer," *Nat. Genet.* 24(3):236-44 (2000)) and in diagnosis to determine appropriate treatment strategies (see, for example, Sgroi *et al.*, "*In vivo* Gene Expression Profile Analysis of Human Breast Cancer Progression," *Cancer Res.* 59(22):5656-61 (1999)).

In microarray-based gene expression screens of pharmacological drug candidates upon cells, each probe provides specific useful data. In particular, it should be appreciated that even those probes that show no change in expression are as informative as those that do change, serving, in essence, as negative controls.

For example, where gene expression analysis is used to assess toxicity of chemical agents on cells, the failure of the agent to change a gene's expression level is evidence that the drug likely does not affect the pathway of which the gene's expressed protein is a part.

Analogously, where gene expression analysis is used to assess side effects of pharmacological agents – whether in lead compound discovery or in subsequent screening of lead compound derivatives – the inability of the agent to alter a gene's expression level is evidence that the drug does not affect the pathway of which the gene's expressed protein is a part.

WO 99/58720 provides methods for quantifying the relatedness of a first and second gene expression profile and for ordering the relatedness of a plurality of gene expression profiles. The methods so described permit useful information to be extracted from a greater percentage of the individual gene expression measurements from a microarray than methods previously used in the art.

Other uses of microarrays are described in Gerhold *et al.*, *Trends Biochem. Sci.* 24(5):168-173 (1999) and Zweiger, *Trends Biotechnol.* 17(11):429-436 (1999); Schena et al.

The invention particularly provides genome-derived single-exon probes known to be expressed in brain.

The individual single exon probes can be provided in the form of substantially isolated and purified nucleic acid, typically, but not necessarily, in a quantity sufficient to perform a hybridization reaction.

Such nucleic acid can be in any form directly hybridizable to the message that contains the probe's ORF, such as double stranded DNA, single-stranded DNA complementary to the message, single-stranded RNA complementary to the message, or chimeric DNA/RNA molecules so hybridizable. The nucleic acid can alternatively or additionally include either nonnative nucleotides, alternative internucleotide linkages, or both, so long as complementary binding can be obtained. For example, probes can include phosphorothioates, methylphosphonates, morpholino analogs, and peptide nucleic acids (PNA), as are

described, for example, in U.S. Patent Nos. 5,142,047; 5,235,033; 5,166,315; 5,217,866; 5,184,444; 5,861,250.

Usefully, however, such probes are provided in a form and quantity suitable for amplification, where the amplified product is thereafter to be used in the hybridization reactions that probe gene expression. Typically, such probes are provided in a form and quantity suitable for amplification by PCR or by other well known amplification technique. One such technique additional to PCR is rolling circle amplification, as is described, *inter alia*, in U.S. Patent Nos. 5,854,033 and 5,714,320 and international patent publications WO 97/19193 and WO 00/15779. As is well understood, where the probes are to be provided in a form suitable for amplification, the range of nucleic acid analogues and/or internucleotide linkages will be constrained by the requirements and nature of the amplification enzyme.

Where the probe is to be provided in form suitable for amplification, the quantity need not be sufficient for direct hybridization for gene expression analysis, and need be sufficient only to function as an amplification template, typically at least about 1, 10 or 100 pg or more.

Each discrete amplifiable probe can also be packaged with amplification primers, either in a single composition that comprises probe template and primers, or in a kit that comprises such primers separately packaged therefrom. As earlier mentioned, the ORF-specific 5' primers used for genomic amplification can have a first common sequence added thereto, and the ORF-specific 3' primers used for genomic amplification can have a second, different, common sequence added thereto, thus permitting, in this embodiment, the use of a single set of 5' and 3' primers to amplify any one of the probes. The probe composition and/or kit can also include buffers, enzyme,

etc., required to effect amplification.

As mentioned earlier, when intended for use on a genome-derived single exon microarray of the present invention, the genome-derived single exon probes of the present invention will typically average at least about 100, 200, 300, 400 or 500 bp in length, including (and typically, but not necessarily centered about) the ORF. Furthermore, when intended for use on a genome-derived single exon microarray of the present invention, the genome-derived single exon probes of the present invention will typically not contain a detectable label.

When intended for use in solution phase hybridization, however – that is, for use in a hybridization reaction in which the probe is not first bound to a support substrate (although the target may indeed be so bound) – length constraints that are imposed in microarray-based hybridization approaches will be relaxed, and such probes will typically be labeled.

In such case, the only functional constraint that dictates the minimum size of such probe is that each such probe must be capable of specifically identifying in a hybridization reaction the exon from which it is drawn. In theory, a probe of as little as 17 nucleotides is capable of uniquely identifying its cognate sequence in the human genome. For hybridization to expressed message – a subset of target sequence that is much reduced in complexity as compared to genomic sequence – even fewer nucleotides are required for specificity.

Therefore, the probes of the present invention can include as few as 20, 25 or 50 bp or ORF, or more. In particular embodiments, the ORF sequences are given in SEQ ID NOS. 12,822 - 25,434, respectively, for probe SEQ ID NOS. 1 - 12,821. The minimum amount of ORF required to be included in the probe of the present invention in order to provide specific signal in either solution phase or

microarray-based hybridizations can readily be determined for each of ORF SEQ ID NOS. 12,822 - 25,434 individually by routine experimentation using standard high stringency conditions.

Such high stringency conditions are described, *inter alia*, in Ausubel et al. and Maniatis *et al.* For microarray-based hybridization, standard high stringency conditions can usefully be 50% formamide, 5X SSC, 0.2 μg/μl poly(dA), 0.2 μg/μl human $c_o t1$ DNA, and 0.5 % SDS, in a humid oven at 42°C overnight, followed by successive washes of the microarray in 1X SSC, 0.2% SDS at 55°C for 5 minutes, and then 0.1X SSC, 0.2% SDS, at 55°C for 20 minutes. For solution phase hybridization, standard high stringency conditions can usefully be aqueous hybridization at 65°C in 6X SSC. Lower stringency conditions, suitable for cross-hybridization to mRNA encoding structurally- and functionally-related proteins, can usefully be the same as the high stringency conditions but with reduction in temperature for hybridization and washing to room temperature (approximately 25°C).

When intended for use in solution phase hybridization, the maximum size of the single exon probes of the present invention is dictated by the proximity of other expressed exons in genomic DNA: although each single exon probe can include intergenic and/or intronic material contiguous to the ORF in the human genome, each probe of the present invention will include portions of only one expressed exon.

Thus, each single exon probe will include no more than about 25 kb of contiguous genomic sequence, more typically no more than about 20 kb of contiguous genomic sequence, more usually no more than about 15 kb, even more usually no more than about 10 kb. Usually, probes that are maximally about 5 kb will be used, more typically no more than about 3 kb.

It will be appreciated that the Sequence Listing appended hereto presents, by convention, only that strand of the probe and ORF sequence that can be directly translated reading from 5' to 3' end. As would be well understood by one of skill in the art, single stranded probes must be complementary in sequence to the ORF as present in an mRNA; it is well within the skill in the art to determine such complementary sequence. It will further be understood that double stranded probes can be used in both solution-phase hybridization and microarray-based hybridization if suitably denatured.

Thus, it is an aspect of the present invention to provide single-stranded nucleic acid probes that have sequence complementary to those described herein above and below, and double-stranded probes one strand of which has sequence complementary to the probes described herein.

The probes can, but need not, contain intergenic and/or intronic material that flanks the ORF, on one or both sides, in the same linear relationship to the ORF that the intergenic and/or intronic material bears to the ORF in genomic DNA. The probes do not, however, contain nucleic acid derived from more than one expressed ORF.

And when intended for use in solution hybridization, the probes of the present invention can usefully have detectable labels. Nucleic acid labels are well known in the art, and include, *inter alia*, radioactive labels, such as ^{3}H, ^{32}P, ^{33}P, ^{35}S, ^{125}I, ^{131}I; fluorescent labels, such as Cy3, Cy5, Cy5.5, Cy7, SYBR®

Green and other labels described in Haugland, *Handbook of Fluorescent Probes and Research Chemicals*, 7th ed., Molecular Probes Inc., Eugene, OR (2000), or fluorescence resonance energy transfer tandem conjugates thereof; labels suitable for chemiluminescent and/or enhanced chemiluminescent detection; labels suitable for ESR and NMR detection; and labels that include one member

of a specific binding pair, such as biotin, digoxigenin, or the like.

The probes, either in quantity sufficient for hybridization or sufficient for amplification, can be provided in individual vials or containers.

Alternatively, such probes can usefully be packaged as a plurality of such individual genome-derived single exon probes.

When provided as a collection of plural individual probes, the probes are typically made available in amplifiable form in a spatially-addressable ordered set, typically one per well of a microtiter dish. Although a 96 well microtiter plate can be used, greater efficiency is obtained using higher density arrays.

If, as earlier mentioned, the ORF-specific 5' primers used for genomic amplification had a first common sequence added thereto, and the ORF-specific 3' primers used for genomic amplification had a second, different, common sequence added thereto, a single set of 5' and 3' primers can be used to amplify all of the probes from the amplifiable ordered set.

Such collections of genome-derived single exon probes can usefully include a plurality of probes chosen for the common attribute of expression in the human brain.

In such defined subsets, typically at least 50, 60, 75, 80, 85, 90 or 95% or more of the probes will be chosen by their expression in the defined tissue or cell type.

The single exon probes of the present invention, as well as fragments of the single exon probes comprising selectively hybridizable portions of the probe ORF, can be used to obtain the full length cDNA that includes the ORF by (i) screening of cDNA libraries; (ii) rapid amplification of cDNA ends ("RACE"); or (iii) other conventional means, as are described, *inter alia*, in

Ausubel et al. and Maniatis *et al.*

It is another aspect of the present invention to provide genome-derived single exon nucleic acid microarrays useful for gene expression analysis, where the term "microarray" has the meaning given in the definitional section of this description, *supra*.

The invention particularly provides genome-derived single-exon nucleic acid microarrays comprising a plurality of probes known to be expressed in human brain. In preferred embodiments, the present invention provides human genome-derived single exon microarrays comprising a plurality of probes drawn from the group consisting of SEQ ID NOS.: 1 - 12,821.

When used for gene expression analysis, the genome-derived single exon microarrays provide greater physical informational density than do the genome-derived single exon microarrays that have lower percentages of probes known to be expressed commonly in the tested tissue. At a fixed probe density, for example, a given microarray surface area of the defined subset genome-derived single exon microarray can yield a greater number of expression measurements. Alternatively, at a given probe density, the same number of expression measurements can be obtained from a smaller substrate surface area. Alternatively, at a fixed probe density and fixed surface area, probes can be provided redundantly, providing greater reliability in signal measurement for any given probe. Furthermore, with a higher percentage of probes known to be expressed in the assayed tissue, the dynamic range of the detection means can be adjusted to reveal finer levels discrimination among the levels of expression.

Although particularly described with respect to their utility as probes of gene expression, particularly as probes to be included on a genome-derived single exon microarray, each of the nucleic acids having SEQ ID NOS.: 1

- 12,821 contains an open-reading frame, set forth respectively in SEQ ID NOS.: 12,822 - 25,434, that encodes a protein domain. Thus, each of SEQ ID NOS. 1 - 12,821 can be used, or that portion thereof in SEQ ID NOS. 12,822 - 25,434 used, to express a protein domain by standard *in vitro* recombinant techniques. See Ausubel et al. and Maniatis *et al.*

Additionally, kits are available commercially that readily permit such nucleic acids to be expressed as protein in bacterial cells, insect cells, or mammalian cells, as desired (*e.g.*, HAT™ Protein Expression & Purification System, ClonTech Laboratories, Palo Alto, CA; Adeno-X™ Expression System, ClonTech Laboratories, Palo Alto, CA; Protein Fusion & Purification (pMAL™) System, New England Biolabs, Beverley, MA)

Furthermore, shorter peptides can be chemically synthesized using commercial peptide synthesizing equipment and well known techniques. Procedures are described, *inter alia*, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series, (Paper)), Oxford Univ. Press (March 2000) (ISBN: 0199637245); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7) , Oxford Univ. Press (August 1992) (ISBN: 0198556683); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (December 1993) (ISBN: 0387564314).

It is, therefore, another aspect of the invention to provide peptides comprising an amino acid sequence translated from SEQ ID NOS.: 12,822 - 25,434. Such amino acid sequences are set out in SEQ ID NOS: 25,435 - 37,811. Any such recombinantly-expressed or synthesized peptide of at least 8, and preferably at least about 15, amino acids, can be conjugated to a carrier protein and used to generate antibody that recognizes the peptide. Thus, it is a further aspect of the invention to provide peptides that

have at least 8, preferably at least 15, consecutive amino acids.

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLE 1

Preparation of Single Exon Microarrays from ORFs Predicted in Human Genomic Sequence

Bioinformatics Results

All human BAC sequences in fewer than 10 pieces that had been accessioned in a five month period immediately preceding this study were downloaded from GenBank. This corresponds to ~2200 clones, totaling ~350 MB of sequence, or approximately 10% of the human genome.

After masking repetitive elements using the program CROSS_MATCH, the sequence was analyzed for open reading frames using three separate gene finding programs. The three programs predict genes using independent algorithmic methods developed on independent training sets: GRAIL uses a neural network, GENEFINDER uses a hidden Markoff model, and DICTION, a program proprietary to Genetics Institute, operates according to a different heuristic. The results of all three programs were used to create a prediction matrix across the segment of genomic DNA.

The three gene finding programs yielded a range of results. GRAIL identified the greatest percentage of genomic sequence as putative coding region, 2% of the data analyzed. GENEFINDER was second, calling 1%, and DICTION yielded the least putative coding region, with 0.8% of genomic sequence called as coding region.

The consensus data were as follows. GRAIL and GENEFINDER agreed on 0.7% of genomic sequence, GRAIL and

DICTION agreed on 0.5% of genomic sequence, and the three programs together agreed on 0.25% of the data analyzed. That is, 0.25% of the genomic sequence was identified by all three of the programs as containing putative coding region.

ORFs predicted by any two of the three programs ("consensus ORFs") were assorted into "gene bins" using two criteria: (1) any 7 consecutive exons within a 25 kb window were placed together in a bin as likely contributing to a single gene, and (2) all ORFs within a 25 kb window were placed together in a bin as likely contributing to a single gene if fewer than 7 exons were found within the 25 kb window.

PCR

The largest ORF from each gene bin that did not span repetitive sequence was then chosen for amplification, as were all consensus ORFs longer than 500 bp. This method approximated one exon per gene; however, a number of genes were found to be represented by multiple elements.

Previously, we had determined that DNA fragments fewer than 250 bp in length do not bind well to the amino-modified glass surface of the slides used as support substrate for construction of microarrays; therefore, amplicons were designed in the present experiments to approximate 500 bp in length.

Accordingly, after selecting the largest ORF per gene bin, a 500 bp fragment of sequence centered on the ORF was passed to the primer picking software, PRIMER3 (available online for use at http://www-genome.wi.mit.edu/cgi-bin/primer/). A first additional sequence was commonly added to each ORF-unique 5' primer, and a second, different, additional sequence was commonly added to each ORF-unique 3' primer, to permit subsequent reamplification of the amplicon using a single

set of "universal" 5' and 3' primers, thus immortalizing the amplicon. The addition of universal priming sequences also facilitates sequence verification, and can be used to add a cloning site should some ORFs be found to warrant further study.

The ORFs were then PCR amplified from genomic DNA, verified on agarose gels, and sequenced using the universal primers to validate the identity of the amplicon to be spotted in the microarray.

Primers were supplied by Operon Technologies (Alameda, CA). PCR amplification was performed by standard techniques using human genomic DNA (Clontech, Palo Alto, CA) as template. Each PCR product was verified by SYBR® green (Molecular Probes, Inc., Eugene, OR) staining of agarose gels, with subsequent imaging by Fluorimager (Molecular Dynamics, Inc., Sunnyvale, CA). PCR amplification was classified as successful if a single band appeared.

The success rate for amplifying ORFs of interest directly from genomic DNA using PCR was approximately 75%. FIG. 5 graphs the distribution of predicted ORF (exon) length and distribution of amplified PCR products, with ORF length shown in red and PCR product length shown in blue (which may appear black in the figure). Although the range of ORF sizes is readily seen to extend to beyond 900 bp, the mean predicted exon size was only 229 bp, with a median size of 150 bp (n=9498). With an average amplicon size of 475 ± 25 bp, approximately 50% of the average PCR amplification product contained predicted coding region, with the remaining 50% of the amplicon containing either intron, intergenic sequence, or both.

Using a strategy predicated on amplifying about 500 bp, it was found that long exons had a higher PCR failure rate. To address this, the bioinformatics process was adjusted to amplify 1000, 1500 or 2000 bp fragments

from exons larger than 500 bp. This improved the rate of successful amplification of exons exceeding 500 bp, constituting about 9.2% of the exons predicted by the gene finding algorithms.

Approximately 75% of the probes disposed on the array (90% of those that successfully PCR amplified) were sequence-verified by sequencing in both the forward and reverse direction using MegaBACE sequencer (Molecular Dynamics, Inc., Sunnyvale, CA), universal primers, and standard protocols.

Some genomic clones (BACs) yielded very poor PCR and sequencing results. The reasons for this are unclear, but may be related to the quality of early draft sequence or the inclusion of vector and host contamination in some submitted sequence data.

Although the intronic and intergenic material flanking coding regions could theoretically interfere with hybridization during microarray experiments, subsequent empirical results demonstrated that differential expression ratios were not significantly affected by the presence of noncoding sequence. The variation in exon size was similarly found not to affect differential expression ratios significantly; however, variation in exon size was observed to affect the absolute signal intensity (data not shown).

The 350 MB of genomic DNA was, by the above-described process, reduced to 9750 discrete probes, which were spotted in duplicate onto glass slides using commercially available instrumentation (MicroArray GenII Spotter and/or MicroArray GenIII Spotter, Molecular Dynamics, Inc., Sunnyvale, CA). Each slide additionally included either 16 or 32 *E. coli* genes, the average hybridization signal of which was used as a measure of background biological noise.

Each of the probe sequences was BLASTed against

the human EST data set, the NR data set, and SwissProt GenBank (May 7, 1999 release 2.0.9).

One third of the probe sequences (as amplified) produced an exact match (BLAST Expect ("E") values less than 1 e^{-100}) to either an EST (20% of sequences) or a known mRNA (13% of sequences). A further 22% of the probe sequences showed some homology to a known EST or mRNA (BLAST E values from 1 e^{-5} to 1 e^{-99}). The remaining 45% of the probe sequences showed no significant sequence homology to any expressed, or potentially expressed, sequences present in public databases.

All of the probe sequences (as amplified) were then analyzed for protein similarities with the SwissProt database using BLASTX, Gish *et al.*, *Nature Genet.* 3:266 (1993). The predicted functional breakdowns of the 2/3 of probes identical or homologous to known sequences are presented in Table 1.

Table 1

Function of Predicted ORFs As Deduced From Comparative Sequence Analysis			
Total	V6 chip	V7 chip	Function Predicted from Comparative Sequence Analysis
211	96	115	Receptor
120	43	77	Zinc Finger
30	11	19	Homeobox
25	9	16	Transcription Factor
17	11	7	Transcription
118	57	61	Structural
95	39	56	Kinase
36	18	18	Phosphatase
83	31	52	Ribosomal

45	19	26	Transport
21	17	14	Growth Factor
17	12	5	Cytochrome
50	33	17	Channel

As can be seen, the two most common types of genes were transcription factors and receptors, making up 2.2% and 1.8% of the arrayed elements, respectively.

EXAMPLE 2

Gene Expression Measurements From Genome-Derived Single Exon Microarrays

The two genome-derived single exon microarrays prepared according to Example 1 were hybridized in a series of simultaneous two-color fluorescence experiments to (1) Cy3-labeled cDNA synthesized from message drawn individually from each of brain, heart, liver, fetal liver, placenta, lung, bone marrow, HeLa, BT 474, or HBL 100 cells, and (2) Cy5-labeled cDNA prepared from message pooled from all ten tissues and cell types, as a control in each of the measurements. Hybridization and scanning were carried out using standard protocols and Molecular Dynamics equipment.

Briefly, mRNA samples were bought from commercial sources (Clontech, Palo Alto, CA and Amersham Pharmacia Biotech (APB)). Cy3-dCTP and Cy5-dCTP (both from APB) were incorporated during separate reverse transcriptions of 1 µg of polyA$^+$ mRNA performed using 1 µg oligo(dT)12-18 primer and 2 µg random 9mer primers as follows. After heating to 70°C, the RNA:primer mixture was snap cooled on ice. After snap cooling on ice, added to the RNA to the stated final concentration was: 1X Superscript II buffer, 0.01 M DTT,

100µM dATP, 100 µM dGTP, 100 µM dTTP, 50 µM dCTP, 50 µM Cy3-dCTP or Cy5-dCTP 50 µM, and 200 U Superscript II enzyme. The reaction was incubated for 2 hours at 42°C. After 2 hours, the first strand cDNA was isolated by adding 1 U Ribonuclease H, and incubating for 30 minutes at 37°C. The reaction was then purified using a Qiagen PCR cleanup column, increasing the number of ethanol washes to 5. Probe was eluted using 10 mM Tris pH 8.5.

Using a spectrophotometer, probes were measured for dye incorporation. Volumes of both Cy3 and Cy5 cDNA corresponding to 50 pmoles of each dye were then dried in a Speedvac, resuspended in 30 µl hybridization solution containing 50% formamide, 5X SSC, 0.2 µg/µl poly(dA), 0.2 µg/µl human c_0t1 DNA, and 0.5 % SDS.

Hybridizations were carried out under a coverslip, with the array placed in a humid oven at 42°C overnight. Before scanning, slides were washed in 1X SSC, 0.2% SDS at 55°C for 5 minutes, followed by 0.1X SSC, 0.2% SDS, at 55°C for 20 minutes. Slides were briefly dipped in water and dried thoroughly under a gentle stream of nitrogen.

Slides were scanned using a Molecular Dynamics Gen3 scanner, as described. Schena (ed.), <u>Microarray Biochip: Tools and Technology</u>, Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376).

Although the use of pooled cDNA as a reference permitted the survey of a large number of tissues, it attenuates the measurement of relative gene expression, since every highly expressed gene in the tissue/cell type-specific fluorescence channel will be present to a level of at least 10% in the control channel. Because of this fact, both signal and expression ratios (the latter hereinafter, "expression" or "relative expression") for each probe were normalized using the average ratio or average signal,

respectively, as measured across the whole slide.

Data were accepted for further analysis only when signal was at least three times greater than biological noise, the latter defined by the average signal produced by the *E. coli* control genes.

The relative expression signal for these probes was then plotted as function of tissue or cell type, and is presented in FIG. 6.

FIG. 6 shows the distribution of expression across a panel of ten tissues. The graph shows the number of sequence-verified products that were either not expressed ("0"), expressed in one or more but not all tested tissues ("1" - "9"), and expressed in all tissues tested ("10").

Of 9999 arrayed elements on the two microarrays (including positive and negative controls and "failed" products), 2353 (51%) were expressed in at least one tissue or cell type. Of the gene elements showing significant signal – where expression was scored as "significant" if the normalized Cy3 signal was greater than 1, representing signal 5-fold over biological noise (0.2) – 39% (991) were expressed in all 10 tissues. The next most common class (15%) consisted of gene elements expressed in only a single tissue.

The genes expressed in a single tissue were further analyzed, and the results of the analyses are compiled in FIG. 7.

FIG. 7A is a matrix presenting the expression of all verified sequences that showed expression greater than 3 in at least one tissue. Each clone is represented by a column in the matrix. Each of the 10 tissues assayed is represented by a separate row in the matrix, and relative expression of a clone in that tissue is indicated at the respective node by intensity of green shading, with the intensity legend shown in panel B. The top row of the

matrix ("EST Hit") contains "bioinformatic" rather than "physical" expression data – that is, presents the results returned by query of EST, NR and SwissProt databases using the probe sequence. The legend for "bioinformatic expression" (*i.e.*, degree of homology returned) is presented in panel C. Briefly, white is known, black is novel, with gray depicting nonidentical with significant homology (white: E values < 1e-100; gray: E values from 1e-05 to 1e-99; black: E values > 1e-05).

As FIG. 7 readily shows, heart and brain were demonstrated to have the greatest numbers of genes that were shown to be uniquely expressed in the respective tissue. In brain, 200 uniquely expressed genes were identified; in heart, 150. The remaining tissues gave the following figures for uniquely expressed genes: liver, 100; lung, 70; fetal liver, 150; bone marrow, 75; placenta, 100; HeLa, 50; HBL, 100; and BT474, 50.

It was further observed that there were many more "novel" genes among those that were up-regulated in only one tissue, as compared with those that were down-regulated in only one tissue. In fact, it was found that ORFs whose expression was measurable in only a single of the tested tissues were represented in sequencing databases at a rate of only 11%, whereas 36% of the ORFs whose expression was measurable in 9 of the tissues were present in public databases. As for those ORFs expressed in all ten tissues, fully 45% were present in existing expressed sequence databases. These results are not unexpected, since genes expressed in a greater number of tissues have a higher likelihood of being, and thus of having been, discovered by EST approaches.

<u>Comparison of Signal from Known and Unknown Genes</u>

The normalized signal of the genes found to have high homology to genes present in the GenBank human EST

database were compared to the normalized signal of those genes not found in the GenBank human EST database. The data are shown in FIG. 8.

FIG. 8 shows the normalized Cy3 signal intensity for all sequence-verified products with a BLAST Expect ("E") value of greater than 1e-30 (designated "unknown") upon query of existing EST, NR and SwissProt databases, and shows in blue the normalized Cy3 signal intensity for all sequence-verified products with a BLAST Expect value of less than 1e-30 ("known"). Note that biological background noise has an averaged normalized Cy3 signal intensity of 0.2.

As expected, the most highly expressed of the ORFs were "known" genes. This is not surprising, since very high signal intensity correlates with very commonly-expressed genes, which have a higher likelihood of being found by EST sequence.

However, a significant point is that a large number of even the high expressers were "unknown". Since the genomic approach used to identify genes and to confirm their expression does not bias exons toward either the 3' or 5' end of a gene, many of these high expression genes will not have been detected in an end-sequenced cDNA library.

The significant point is that presence of the gene in an EST database is *not* a prerequisite for incorporation into a genome-derived microarray, and further, that arraying such "unknown" exons can help to assign function to as-yet undiscovered genes.

<u>Verification of Gene Expression</u>

To ascertain the validity of the approach described above to identify genes from raw genomic sequence, expression of two of the probes was assayed using reverse transcriptase polymerase chain reaction (RT PCR)

and northern blot analysis.

Two microarray probes were selected on the basis of exon size, prior sequencing success, and tissue-specific gene expression patterns as measured by the microarray experiments. The primers originally used to amplify the two respective ORFs from genomic DNA were used in RT PCR against a panel of tissue-specific cDNAs (Rapid-Scan gene expression panel 24 human cDNAs) (OriGene Technologies, Inc., Rockville, MD).

Sequence AL079300_1 was shown by microarray hybridization to be present in cardiac tissue, and sequence AL031734_1 was shown by microarray experiment to be present in placental tissue (data not shown). RT-PCR on these two sequences confirmed the tissue-specific gene expression as measured by microarrays, as ascertained by the presence of a correctly sized PCR product from the respective tissue type cDNAs.

Clearly, all microarray results cannot, and indeed should not, be confirmed by independent assay methods, or the high throughput, highly parallel advantages of microarray hybridization assays will be lost. However, in addition to the two RT-PCR results presented above, the observation that 1/3 of the arrayed genes exist in expression databases provides powerful confirmation of the power of our methodology – which combines bioinformatic prediction with expression confirmation using genome-derived single exon microarrays – to identify novel genes from raw genomic data.

To verify that the approach further provides correct characterization of the expression patterns of the identified genes, a detailed analysis was performed of the microarrayed sequences that showed high signal in brain.

For this latter analysis, sequences that showed high (normalized) signal in brain, but which showed very low (normalized) signal (less than 0.5, determined to be

biological noise) in all other tissues, were further studied. There were 82 sequences that fit these criteria, approximately 2% of the arrayed elements. The 10 sequences showing the highest signal in brain in microarray hybridizations are detailed in Table 2, along with assigned function, if known or reasonably predicted.

Table 2

Function of the Most Highly Expressed Genes Expressed Only in Brain				
Microarray Sequence Name	Normalized Signal	Expression Ratio	Homology to EST present in GenBank	Gene Function as described by GenBank
AP000217-1	5.2	+7.7	High	S-100 protein, b-chain, Ca^{2+} binding protein expressed in central nervous system
AP000047-1	2.3		High	Unknown Function
AC006548-9	1.7		High	Similar to mouse membrane glyco-protein M6, expressed in central nervous system

AC007245-5	1.5		High	Similar to amphiphysin, a synaptic vesicle-associated protein. Ref 21
L44140-4	1.2	+2.0	High	Endothelial actin-binding protein found in nonmuscle filamin
AC004689-9	1.2	+3.5	High	Protein Phosphatase PP2A, neuronal/ downregulates activated protein kinases
AL031657-1	1.2	+3.0	High	Unknown function/ Contains the anhyrin motif, a common protein sequence motif
AC009266-2	1.1	+3.7	Low	Low homology to the Synaptotagmin I protein in rat/present at low levels throughout rat brain
AP000086-1	1.0	+2.7	Low	Unknown, very poor homology to collagen

AC004689-3	1.0		High	Protein Phosphatase PP2A, neuronal/ downregulates activated protein kinases

Of the ten sequences studied by these latter confirmatory approaches, eight were previously known. Of these eight, six had previously been reported to be important in the central nervous system or brain. The exon giving the highest signal (AP00217-1) was found to be the gene encoding an S100B Ca^{2+} binding protein, reported in the literature to be highly and uniquely expressed in the central nervous system. Heizmann, *Neurochem. Res.* 9:1097 (1997).

A number of the brain-specific probe sequences (including AC006548-9, AC009266-2) did not have homology to any known human cDNAs in GenBank but did show homology to rat and mouse cDNAs. Sequences AC004689-9 and AC004689-3 were both found to be phosphatases present in neurons (Millward *et al.*, *Trends Biochem. Sci.* 24(5):186-191 (1999)). Two microarray sequences, AP000047-1 and AP000086-1 have unknown function, with AP000086-1 being absent from GenBank. Functionality can now be narrowed down to a role in the central nervous system for both of these genes, showing the power of designing microarrays in this fashion.

Next, the function of the chip sequences with the highest (normalized) signal intensity in brain, regardless of expression in other tissues, was assessed. In this latter analysis, we found expression of many more common genes, since the sequences were not limited to those expressed only in brain. For example, looking at the 20 highest signal intensity spots in brain, 4 were similar to

tubulin (AC00807905; AF146191-2; AC007664-4; AF14191-2), 2 were similar to actin (AL035701-2; AL034402-1), and 6 were found to be homologous to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (AL035604-1; Z86090-1; AC006064-L, AC006064-K; AC035604-3; AC006064-L). These genes are often used as controls or housekeeping genes in microarray experiments of all types.

Other interesting genes highly expressed in brain were a ferritin heavy chain protein, which is reported in the literature to be found in brain and liver (Joshi *et al.*, *J. Neurol. Sci.* 134(Suppl):52-56 (1995)), a result duplicated with the array. Other highly expressed chip sequences included a translation elongation factor 1□ (AC007564-4), a DEAD-box homolog (AL023804-4), and a Y-chromosome RNA-binding motif (Chai *et al.*, *Genomics* 49(2):283-89 (1998))(AC007320-3). A low homology analog (AP00123-1/2) to a gene, DSCR1, thought to be involved in trisomy 21 (Down's syndrome), showed high expression in both brain and heart, in agreement with the literature (Fuentes *et al.*, *Mol. Genet.* 4(10):1935-44 (1995)).

As a further validation of the approach, we selected the BAC AC006064 to be included on the array. This BAC was known to contain the GAPDH gene, and thus could be used as a control for the ORF selection process. The gene finding and exon selection algorithms resulted in choosing 25 exons from BAC AC006064 for spotting onto the array, of which four were drawn from the GAPDH gene. Table 3 shows the comparison of the average expression ratio for the 4 exons from BAC006064 compared with the average expression ratio for 5 different dilutions of a commercially available GAPDH cDNA (Clontech).

Table 3

Comparison of Expression Ratio, for each tissue, of GAPDH		
	AC006064 (n = 4)	Control (n = 5)
Bone Marrow	-1.81 ± 0.11	-1.85 ± 0.08
Brain	-1.41 ± 0.11	-1.17 ± 0.05
BT474	1.85 ± 0.09	1.66 ± 0.12
Fetal Liver	-1.62 ± 0.07	-1.41 ± 0.05
HBL100	1.32 ± 0.05	2.64 ± 0.12
Heart	1.16 ± 0.09	1.56 ± 0.10
HeLa	1.11 ±0.06	1.30 ± 0.15
Liver	-1.62 ± 0.22	-2.07 ±
Lung	-4.95 ± 0.93	-3.75 ± 0.21
Placenta	-3.56 ± 0.25	-3.52 ± 0.43

Each tissue shows excellent agreement between the experimentally chosen exons and the control, again demonstrating the validity of the present exon mining approach. In addition, the data also show the variability of expression of GAPDH within tissues, calling into question its classification as a housekeeping gene and utility as a housekeeping control in microarray experiments.

EXAMPLE 3

Representation of Sequence and Expression Data as a "Mondrian"

For each genomic clone processed for microarray as above-described, a plethora of information was accumulated, including full clone sequence, probe sequence within the clone, results of each of the three gene finding programs, EST information associated with the probe

sequences, and microarray signal and expression for multiple tissues, challenging our ability to display the information.

Accordingly, we devised a new tool for visual display of the sequence with its attendant annotation which, in deference to its visual similarity to the paintings of Piet Mondrian, is hereinafter termed a "Mondrian". FIGS. 3 and 4 present the key to the information presented on a Mondrian.

FIG. 9 presents a Mondrian of BAC AC008172 (bases 25,000 to 130,000 shown), containing the carbamyl phosphate synthetase gene (AF154830.1). Purple background within the region shown as field 81 in FIG. 3 indicates all 37 known exons for this gene.

As can be seen, GRAIL II successfully identified 27 of the known exons (73%), GENEFINDER successfully identified 37 of the known exons (100%), while DICTION identified 7 of the known exons (19%).

Seven of the predicted exons were selected for physical assay, of which 5 successfully amplified by PCR and were sequenced. These five exons were all found to be from the same gene, the carbamyl phosphate synthetase gene (AF154830.1).

The five exons were arrayed, and gene expression measured across 10 tissues. As is readily seen in the Mondrian, the five chip sequences on the array show identical expression patterns, elegantly demonstrating the reproducibility of the system.

FIG. 10 is a Mondrian of BAC AL049839. We selected 12 exons from this BAC, of which 10 successfully sequenced, which were found to form between 5 and 6 genes. Interestingly, 4 of the genes on this BAC are protease inhibitors. Again, these data elegantly show that exons selected from the same gene show the same expression patterns, depicted below the red line. From this figure,

it is clear that our ability to find known genes is very good. A novel gene is also found from 86.6 kb to 88.6 kb, upon which all the exon finding programs agree. We are confident we have two exons from a single gene since they show the same expression patterns and the exons are proximal to each other. Backgrounds in the following colors indicate a known gene (top to bottom): red = kallistatin protease inhibitor (P29622); purple = plasma serine protease inhibitor (P05154); turquoise = α1 anti-chymotrypsin (P01011); mauve = 40S ribosomal protein (P08865). Note that chip sequence 8 and 12 did not sequence verify.

EXAMPLE 4

Genome-Derived Single Exon Probes Useful For Measuring Human Gene Expression

The protocols set forth in Examples 1 and 2, *supra*, were applied to additional human genomic sequence as it became newly available in GenBank to identify unique exons in the human genome that could be shown to be expressed at significant levels in brain tissue.

These unique exons are within longer probe sequences. Each probe was completely sequenced on both strands prior to its use on a genome-derived single exon microarray; sequencing confirms the exact chemical structure of each probe. An added benefit of sequencing is that it placed us in possession of a set of single base-incremented fragments of the sequenced nucleic acid, starting from the sequencing primer 3' OH. (Since the single exon probes were first obtained by PCR amplification from genomic DNA, we were of course additionally in possession of an even larger set of single base incremented fragments of each of the 12,821 single exon probes, each

fragment corresponding to an extension product from one of the two amplification primers.)

The structures of the 12,821 unique single exon probes are clearly presented in the Sequence Listing as SEQ ID Nos.: 1 - 12,821. The 16 nt 5' primer sequence and 16 nt 3' primer sequence present on the amplicon are not included in the sequence listing. The sequences of the exons present within each of these probes is presented in the Sequence Listing as SEQ ID NOs.: 12,822 - 25,434, respectively. It will be noted that some amplicons have more than one exon, some exons are contained in more than one amplicon.

As detailed in Example 2, expression was demonstrated by disposing the amplicons as single exon probes on nucleic acid microarrays and then performing two-color fluorescent hybridization analysis; significant expression is based on a statistical confidence that the signal is significantly greater than negative biological control spots. The negative biological control is formed from spotted DNA sequences from a different species. Here, 32 sequences from E.Coli were spotted in duplicate to give a total of 64 spots.

For each hybridisation (each slide, each colour) the median value of the signal from all of the spots is determined. The normalised signal value is the arithmetic mean of the signal from duplicate spots divided by the population median.

Control spots are eliminated if there is more that a five-fold difference between each one of the duplicate spots raw signals.

The median of the signal from the remaining control spots is calculated and all subsequent calculations are done with normalised signals.

Control spots having a signal of greater than median + 2.4 (the value 2.4 is roughly 12 times the

observed standard deviation of control spot populations) are eliminated. Spots with such high signals are considered to be "outliers".

The mean and standard deviation of the modified control spot populations are calculated.

The mean + 3x the standard deviation (mean + (3*SD)) is used as the signal threshold qualifier for that particular hybridisation. Thus, individual thresholds are determined for each channel and each hybridisation.

This means that, assuming that the data is distributed normally, there is a 99% confidence that any signal exceeding the threshold is significant.

The probes and their expression data are presented in Table 4, set forth respectively in Example 5. Example 5 presents the subset of probes that is significantly expressed in the human heart and thus presents the subset of probes that was recognized to be useful for measuring expression of their cognate genes in human brain tissue.

The sequence of each of the exon probes identified by SEQ ID NOS.: 12,822 - 25,434 was individually used as a BLAST (or, for SWISSPROT, BLASTX) query to identify the most similar sequence in each of dbEST, SwissProt (BLASTX), and NR divisions of GenBank. Because the query sequences are themselves derived from genomic sequence in GenBank, only nongenomic hits from NR were scored.

The smallest in value of the BLAST (or BLASTX) expect ("E") scores for each query sequence across the three database divisions was used as a measure of the "expression novelty" of the probe's ORF. Table 4 is sorted in descending order based on this measure, reported as "Most Similar (top) Hit BLAST E Value". Those sequences for which no "Hit E Value" is listed are those exons which were found to have no similar sequences.

As sorted, Table 4 thus lists its respective probes (by "AMPLICON SEQ ID NO.:" and additionally by the SEQ ID NO:. of the exon contained within the probe:"EXON SEQ ID NO.:") from least similar to sequences known to be expressed (*i.e.*, highest BLAST E value), at the beginning of the table, to most similar to sequences known to be expressed (*i.e.*, lowest BLAST E value), at the bottom of the table.

Table 4 further provides, for each listed probe, the accession number of the database sequence that yielded the "Most Similar (top) Hit BLAST E Value", along with the name of the database in which the database sequence is found ("Top Hit Database Source").

Table 4 further provides SEQ ID NOS. corresponding to the predicted amino acid sequences where they have been determined for the probe and exon nucleotide sequences. These are set out as PEPTIDE SEQ ID NOS.:. The peptide sequences for a given exon are predicted as follows: Since each chip exon is a consensus sequence drawn from predictions from various exon finding programs (i.e. Grail, GeneFinder and GenScan), the multiple initial ORFs are first determined in a uniform way according to each prediction. In particular, the reading frame for predicting the first amino acid in the peptide sequence always starts with the first base of any codon and ends with the last base of non-termination codon. Next, for each strand of the exon, initial ORFs are merged into one or more final ORFs in an exhaustive process based on the following criteria: 1) the merging ORFs must be overlapping, and 2) the merging ORFs must be in the same frame.

The Sequence Listing, which is a superset of all of the data presented in Table 4, further includes, for each probe, the most similar hit, with accession number and BLAST E value, from the each of the three queried databases.

Table 4 further lists, for each probe, a portion of the descriptor for the top hit ("Top Hit Descriptor") as provided in the sequence database. For those ORFs that are similar in sequence, but nonidentical to known sequences (*e.g.*, those with BLAST E values between about 1e-05 and 1e-100), the descriptor reveals the likely function of the protein encoded by the probe's ORF.

Using BLAST E value cutoffs of 1e-05 (*i.e.*, 1 x 10^{-5}) and 1e-100 (*i.e.*, 1 x 10^{-100}) as evidence of similarity to sequences known to be expressed is of course arbitrary: in Example 2, *supra*, a BLAST E value of 1e-30 was used as the boundary when only two classes were to be defined for analysis (unknown, >1e-30; known <1e-30) (see also FIG. 8). Furthermore, even when the "Most Similar (Top) Hit BLAST E Value" is low, *e.g.*, less than about 1e-100 – which is probative evidence that the query sequence has previously been shown to be expressed – the top hit is highly unlikely exactly to match the probe sequence.

First, such expression entries typically will not have the intronic and/or intergenic sequence present within the single exon probes listed in the Table. Second, even the ORF itself is unlikely in such cases to be present identically in the databases, since most of the EST and mRNA clones in existing databases include multiple exons, without any indication of the location of exon boundaries.

As noted, the data presented in Table 4 represent a proper subset of the data present within the attached sequence listing. For each amplicon probe (SEQ ID NOs.: 1 - 12,821) and probe exon (SEQ ID NOs.: 12,822 - 25,434, respectively), the sequence listing further provides, through iterated annotation fields <220> and <223>:

(a) the accession number of the BAC from which the sequence was derived ("MAP TO"), thus providing a link to the chromosomal map location and other information about the genomic milieu of the probe sequence;

(b) the most similar sequence provided by BLAST query of the EST database, with accession number and BLAST E value for the "hit";

(c) the most similar sequence provided by BLAST query of the GenBank NR database, with accession number and BLAST E value for the "hit"; and

(d) the most similar sequence provided by BLASTX query of the SWISSPROT database, with accession number and BLAST E value for the "hit".

EXAMPLE 5

Genome-Derived Single Exon Probes Useful For Measuring Expression of Genes in Human Brain

Table 4 (536 pages) presents expression, homology, and functional information for the genome-derived single exon probes that are expressed significantly in human brain.

CLAIMS

1. A spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human brain comprising a plurality single exon nucleic probes, said probes comprising any one of the nucleotide sequences set out in SEQ ID NOs: 1 - 12,821 or a complementary sequence, or a portion of such a sequence.

2. A spatially-addressable set of single exon nucleic acid probes as claimed in claim 1 wherein each of said plurality of probes is separately and addressably amplifiable.

3. A spatially-addressable set of single exon nucleic acid probes as claimed in claim 1 wherein each of said plurality of probes is separately and addressably isolatable from said plurality.

4. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 3 wherein said probes comprise any one of the nucleotide sequences set out in SEQ ID NOS.: 12,822 - 25,434.

5. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 4, wherein each of said plurality of probes is amplifiable using at least one common primer.

6. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 5 wherein the set comprises between 50 - 20,000 single exon nucleic acid probes.

7. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 6, wherein the

average length of the single exon nucleic acid probes is between 200 and 500 bp.

8. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 7, wherein at least 50% of said single exon nucleic acid probes lack prokaryotic and bacteriophage vector sequence.

9. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 8, wherein at least 50% of said single exon nucleic acid probes lack homopolymeric stretches of A or T.

10. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 - 9 characterised in that said set of probes is addressably disposed upon a substrate.

11. A spatially-addressable set of single exon nucleic acid probes as claimed in claim 10 wherein said substrate is selected from glass, amorphous silicon, crystalline silicon and plastic.

12. A microarray comprising a spatially addressable set of single exon nucleic acid probes as claimed in any of claims 1 - 11.

13. A single exon nucleic acid probe for measuring human gene expression in a sample derived from human brain comprising a nucleotide sequence as set out in any of SEQ ID NOs.: 1 - 12,821 or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid molecule expressed in the human brain.

14. A single exon nucleic acid probe as claimed in claim 13 comprising a nucleotide sequence as set out in any of SEQ ID NOs.: 12,822 - 25,434 or a complementary sequence or a fragment thereof.

15. A single exon nucleic acid probe for measuring human gene expression in a sample derived from human brain which is a nucleic acid molecule having a sequence encoding a peptide comprising a peptide sequence as set out in any of SEQ ID NOs.: 25,435 - 37,811, or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid expressed in the human brain.

16. A single exon nucleic acid probe as claimed in any one of claims 13 to 15 wherein said single exon nucleic acid probe comprises between 15 and 25 contiguous nucleotides of said SEQ ID NO.

17. A single exon nucleic acid probe as claimed in any one of claims 13 to 15, wherein said probe is between 3 - 25 kb in length.

18. A single exon nucleic acid probe as claimed in any one of claims 13 - 17, wherein said probe is DNA, RNA or PNA.

19. A single exon nucleic acid probe as claimed in any one of claims 13 - 18, wherein said probe is detectably labeled.

20. A single exon nucleic acid probe as claimed in any one of claims 13 - 19, wherein said probe lacks prokaryotic and bacteriophage vector sequence.

21. A single exon nucleic acid probe as claimed in any one of claims 13 - 20, wherein said probe lacks homopolymeric

stretches of A or T.

22. A method of measuring gene expression in a sample derived from human brain, comprising:

contacting the microarray of claim 12, with a first collection of detectably labeled nucleic acids, said first collection of nucleic acids derived from mRNA of human brain; and then

measuring the label detectably bound to each probe of said microarray.

23. A method of identifying exons in a eukaryotic genome, comprising:

algorithmically predicting at least one exon from genomic sequence of said eukaryote; and then

detecting specific hybridization of detectably labeled nucleic acids to a single exon probe,

wherein said detectably labeled nucleic acids are derived from mRNA from the brain of said eukaryote, said probe is a single exon probe having a fragment identical in sequence to, or complementary in sequence to, said predicted exon, said probe is included within a microarray according to claim 12, and said fragment is selectively hybridizable at high stringency.

24. A method of assigning exons to a single gene, comprising:

identifying a plurality of exons from genomic sequence according to the method of claim 23; and then

measuring the expression of each of said exons in a plurality of tissues and/or cell types using hybridization to single exon microarrays having a probe with said exon,

wherein a common pattern of expression of said exons in

said plurality of tissues and/or cell types indicates that the exons should be assigned to a single gene.

25. A nucleic acid sequence as set out in any of SEQ ID NOs: 1 - 25,434 which encodes a peptide.

26. A peptide encoded by a sequence as set out in any of SEQ ID Nos: 1 - 25,434.

27. A peptide comprising a sequence as set out in any of SEQ ID NOs: 25,435 - 37,811.

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
437	13223	25868	6.47				
869	13638	26308	15.92				
1022	13782		2.15				
1279	14029	26698	10.88				
1488	14235	26921	1.22				
1488	14235	26922	1.22				
1609	14355	27044	3.19				
1633	14379	27066	6.1				
1718	14461	27160	3.31				
1743	14485	27184	1.44				
1750	14492	27192	6.78				
1884	14621	27331	1.44				
1971	14707	27425	2.14				
2162	14892	27627	2.7				
2277	15003	27743	2.91				
2578	15292	28028	1				
2578	15292	28029	1				
3181	15944	28595	2.83				
3442	16198	28848	1.42				
3505	16261	28915	12.04				
3549	16304		1				
3649	16402	29042	1.67				
3928	16678		1.03				
4173	16913	29543	1.52				
4230	16971	29595	6.4				
4248	16989	29613	0.97				
4248	16989	29614	0.97				
4303	17042		1.07				
4361	17099	29734	0.76				
4784	17516	30138	0.99				
4983	17706	30310	5.38				
4995	17718	30323	1.3				
5176	17985	30500	1.57				
5176	17985	30501	1.57				

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5336	18139		4.3				
5510	18308		6.14				
5593	18139		3.97				
5648	18443	31356	0.6				
5654	18449	31362	3.28				
5932	25082	31673	1.62				
5958	18740	31699	1.76				
6322	19092		1.27				
6454	19222	32220	1.1				
6454	19222	32221	1.1				
7025	19717	32774	1				
7025	19717	32775	1				
7311	19994	33071	1.76				
7311	19994	33072	1.76				
7712	20376		0.61				
7960	20655	33780	1.4				
8384	21077	34214	1.49				
8759	21451	34598	0.59				
8759	21451	34599	0.59				
9434	22112	35287	2.67				
9666	22318	35515	0.77				
9782	22433	35638	1.24				
9922	22570	35767	0.94				
10328	22975	36194	0.62				
10328	22975	36195	0.62				
10582	23277		2.53				
10749	25131	36679	1.34				
10952	23629		2.2				
11030	23701	36968	1.84				
11332	24023	37328	2.02				
11485	24086		2.47				
12313	24735		1.52				
12609	24916	31006	2.36				
5961	18743	31703	17.79	9.9E+00	AJ239028.1	NT	Homo sapiens LSS gene, partial, exons 15, 16, 17 and 18

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7905	20600	33730	1.74	9.8E+00	U32716.1	NT	Haemophilus influenzae Rd section 31 of 163 of the complete genome
9643	22295	35489	0.44	9.8E+00	Y18930.1	NT	Sulfolobus solfataricus 281 kb genomic DNA fragment, strain P2
9643	22295	35490	0.44	9.8E+00	Y18930.1	NT	Sulfolobus solfataricus 281 kb genomic DNA fragment, strain P2
6901	19639	32684	0.73	9.6E+00	AF065630.1	NT	Gallus gallus ornithine transcarbamylase (OTC) gene, exon 1
6901	19639	32685	0.73	9.6E+00	AF065630.1	NT	Gallus gallus ornithine transcarbamylase (OTC) gene, exon 1
10321	22968	36187	1.17	9.6E+00	AF242432.1	NT	Mus musculus Naip3 gene, exon 1; neuronal apoptosis inhibitory protein 1 (Naip1) and general transcription factor IIH polypeptide 2 (Gtf2h2) genes, complete cds
10321	22968	36188	1.17	9.6E+00	AF242432.1	NT	Mus musculus Naip3 gene, exon 1; neuronal apoptosis inhibitory protein 1 (Naip1) and general transcription factor IIH polypeptide 2 (Gtf2h2) genes, complete cds
2671	15381	28119	1	9.4E+00	L11433.1	NT	Dengue virus type 3 membrane protein (prM/M)/envelope glycoprotein (E) polyprotein mRNA, partial cds
2671	15381	28120	1	9.4E+00	L11433.1	NT	Dengue virus type 3 membrane protein (prM/M)/envelope glycoprotein (E) polyprotein mRNA, partial cds
2924	15690	28334	2.87	9.4E+00	AB043785.1	NT	Mus musculus AT3 gene for antithrombin, complete cds
7997	20692	33820	0.91	9.3E+00	AF130990.1	NT	Homo sapiens ectodysplasin-A receptor protein (EDAR) gene, exons 2, 3, and 4
8901	21592	34733	3.06	9.3E+00	P11210	SWISSPROT	IMMEDIATE-EARLY PROTEIN 1 (IE1) (IMMEDIATE-EARLY PHOSPHOPROTEIN PP89)
5214	18022	30645	2.46	9.1E+00	AF095609.1	NT	Leuciscus cephalus orientalis cytochrome b (cyt b) gene, partial cds; mitochondrial gene for mitochondrial product
5214	18022	30646	2.46	9.1E+00	AF095609.1	NT	Leuciscus cephalus orientalis cytochrome b (cyt b) gene, partial cds; mitochondrial gene for mitochondrial product
9330	21997		0.83	9.0E+00	P09241	SWISSPROT	RHODOPSIN
5945	18727	31685	5.55	8.9E+00	BE971806.1	EST_HUMAN	601651038R1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:3934592 3'
6287	19060	32041	2.28	8.7E+00	AB019788.1	NT	Cynops pyrrhogaster CpTbx3 premature mRNA, partial cds
6287	19060	32042	2.28	8.7E+00	AB019788.1	NT	Cynops pyrrhogaster CpTbx3 premature mRNA, partial cds
430	13216	25861	2.3	8.4E+00	5031804	NT	Homo sapiens insulin receptor substrate 1 (IRS1) mRNA
9355	20426	33545	3.58	8.1E+00	AJ131719.1	NT	Zea mays mRNA for legumain-like protease (see2a)
11122	23791		2	8.0E+00	P41820	SWISSPROT	BREFELDIN A RESISTANCE PROTEIN
8051	20745		0.89	7.6E+00	Z21489.1	NT	African swine fever virus NP1450L gene encoding RNA polymerase largest subunit
7246	19931		1.9	7.5E+00	AL445065.1	NT	Thermoplasma acidophilum complete genome; segment 3/5
8259	20953	34090	1.61	7.5E+00	P35441	SWISSPROT	THROMBOSPONDIN 1 PRECURSOR
8259	20953	34091	1.61	7.5E+00	P35441	SWISSPROT	THROMBOSPONDIN 1 PRECURSOR
5711	18504	31426	2.66	7.4E+00	BF700517.1	EST_HUMAN	602128876F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4285506 5'
8651	21343	34487	2.7	7.4E+00	P04929	SWISSPROT	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR
8651	21343	34488	2.7	7.4E+00	P04929	SWISSPROT	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2977	15743	28390	3.58	7.2E+00	L12051.1	NT	Lycopersicon esculentum Mill. GTPase (SAR2) mRNA, complete cds
2977	15743	28391	3.58	7.2E+00	L12051.1	NT	Lycopersicon esculentum Mill. GTPase (SAR2) mRNA, complete cds
6931	19667	32713	0.71	7.2E+00	BE179090.1	EST_HUMAN	RC0-HT0613-200300-031-e07 HT0613 Homo sapiens cDNA
7049	19740	32800	1.28	7.1E+00	P28166	SWISSPROT	ZINC-FINGER PROTEIN 1 (ZINC-FINGER HOMEODOMAIN PROTEIN 1)
7049	19740	32801	1.28	7.1E+00	P28166	SWISSPROT	ZINC-FINGER PROTEIN 1 (ZINC-FINGER HOMEODOMAIN PROTEIN 1)
9498	22151		8.63	7.1E+00	AL161595.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 91
11359	24047	37350	3.28	7.1E+00	P05850	SWISSPROT	HYPOTHETICAL 17.3 KDA PROTEIN IN MRDA-PHPB INTERGENIC REGION
9882	22532	35729	3.37	7.0E+00	P48610	SWISSPROT	ARGININE KINASE (AK)
11215	23878	37165	1.51	7.0E+00	O22469	SWISSPROT	WD-40 REPEAT PROTEIN MSI3
8181	20875	34011	1.92	6.9E+00	P35679	SWISSPROT	60S RIBOSOMAL PROTEIN L4 (L2)
10249	22897	36107	1.38	6.9E+00	P44834	SWISSPROT	DNA MISMATCH REPAIR PROTEIN MUTS
10267	22915	36125	0.47	6.9E+00	P34226	SWISSPROT	SKT5 PROTEIN
7808	20503	33623	1.53	6.8E+00	W03412.1	EST_HUMAN	za07c11.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:291860 5'
7808	20503	33624	1.53	6.8E+00	W03412.1	EST_HUMAN	za07c11.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:291860 5'
9031	21721		1.29	6.8E+00	P36307	SWISSPROT	OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4) [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8]
10109	22757	35969	3.24	6.8E+00	Q03570	SWISSPROT	HYPOTHETICAL 157.0 KDA PROTEIN C38C10.5 IN CHROMOSOME III
5202	18010		0.72	6.6E+00	Q99028	SWISSPROT	CATECHOL-O-METHYLTRANSFERASE, SOLUBLE FORM (S-COMT)
6450	19218	32216	0.61	6.6E+00	BF672121.1	EST_HUMAN	602152573F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293427 5'
9974	22622	35827	2.36	6.6E+00	Q9ZE07	SWISSPROT	URIDYLATE KINASE (UK) (URIDINE MONOPHOSPHATE KINASE) (UMP KINASE)
9974	22622	35828	2.36	6.6E+00	Q9ZE07	SWISSPROT	URIDYLATE KINASE (UK) (URIDINE MONOPHOSPHATE KINASE) (UMP KINASE)
11073	23743		1.97	6.6E+00	Q10309	SWISSPROT	PROBABLE CATION-TRANSPORTING ATPASE C6C3.05C
9079	21768	34931	7	6.5E+00	P03374	SWISSPROT	ENV POLYPROTEIN [CONTAINS: COAT PROTEIN GP52; COAT PROTEIN GP36]
10203	22851	36067	0.49	6.5E+00	BE866001.1	EST_HUMAN	601678435F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3960969 5'
9642	22294	35488	1.55	6.2E+00	AY010901.1	NT	Schizophyllum commune unknown mRNA
10460	23106	36337	0.5	6.2E+00	6754621	NT	Mus musculus mannosidase 2, alpha B1 (Man2b1), mRNA
6936	19671	32717	1.46	6.0E+00	BE780163.1	EST_HUMAN	601468031F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3871303 5'
9716	22367	35565	0.46	6.0E+00	AP000006.1	NT	Pyrococcus horikoshii OT3 genomic DNA, 1166001-1485000 nt. position (6/7)
10411	23057	36274	0.67	6.0E+00	AE001862.1	NT	Deinococcus radiodurans R1 section 1 of 2 of the complete chromosome 2
10411	23057	36275	0.67	6.0E+00	AE001862.1	NT	Deinococcus radiodurans R1 section 1 of 2 of the complete chromosome 2
6428	19196	32193	7.32	5.9E+00	AF155142.1	NT	Mus musculus mixed lineage kinase 3 (Mlk3) and two pore domain K+ channel subunit (Kcnk6) genes, complete cds
3514	16270		0.99	5.8E+00	7661557	NT	Homo sapiens DESC1 protein (DESC1), mRNA
7061	19752	32816	0.95	5.7E+00	AF302046.1	NT	Mus musculus immunoglobulin scavenger receptor IgSR mRNA, complete cds

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7061	19752	32817	0.95	5.7E+00	AF302046.1	NT	Mus musculus immunoglobulin scavenger receptor IgSR mRNA, complete cds
7468	20142		1.13	5.6E+00	P75080	SWISSPROT	DNA POLYMERASE III, ALPHA CHAIN POLC-TYPE (POLIII)
11456	23223	36458	2.59	5.6E+00	Q55276	SWISSPROT	LYCOPENE BETA CYCLASE
6157	18934	31901	0.69	5.5E+00	P47447	SWISSPROT	HEAT-INDUCIBLE TRANSCRIPTION REPRESSOR HRCA
10678	23369		1.28	5.5E+00	AF175425.1	NT	Mus musculus DNA methyltransferase (Dnmt1) gene, exons 30, 31, and 32
11454	23221	36455	3.09	5.5E+00	P11990	SWISSPROT	PNEUMOLYSIN (THIOL-ACTIVATED CYTOLYSIN)
6830	19492	32514	1.14	5.4E+00	X02212.1	NT	Chicken alpha-cardiac actin gene
6830	19492	32515	1.14	5.4E+00	X02212.1	NT	Chicken alpha-cardiac actin gene
7769	20465		1.54	5.4E+00	Q91062	SWISSPROT	VITELLOGENIN PRECURSOR (VTG) [CONTAINS: LIPOVITELLIN LV-1N; LIPOVITELLIN LV-1C; LIPOVITELLIN LV-2]
8698	21390	34534	0.83	5.4E+00	P40379	SWISSPROT	REP1 PROTEIN
8698	21390	34535	0.83	5.4E+00	P40379	SWISSPROT	REP1 PROTEIN
9936	22584	35784	1.83	5.4E+00	Q17094	SWISSPROT	RHODOPSIN
9936	22584	35785	1.83	5.4E+00	Q17094	SWISSPROT	RHODOPSIN
4734	17466	30102	1.32	5.3E+00	L43126.1	NT	Bovine immunodeficiency-like virus surface envelope gene, 5' end of cds
7978	20673		3.23	5.3E+00	P54098	SWISSPROT	DNA POLYMERASE GAMMA (MITOCHONDRIAL DNA POLYMERASE CATALYTIC SUBUNIT)
8882	21573		0.49	5.3E+00	AB034990.1	NT	Homo sapiens HERPUD1 gene for stress protein Herp, complete cds
11628	24225	37548	3.2	5.3E+00	Q27905	SWISSPROT	PROBABLE ANTIBACTERIAL PEPTIDE POLYPROTEIN PRECURSOR
5377	18177		0.91	5.2E+00	BE184840.1	EST_HUMAN	QV4-HT0691-270400-186-f09 HT0691 Homo sapiens cDNA
10271	22919		0.95	5.2E+00	AF248070.1	NT	Drosophila orientacea R1B retrotransposable element reverse transcriptase gene, partial cds
11150	23817		2	5.2E+00	Q10136	SWISSPROT	HYPOTHETICAL 61.1 KD PROTEIN C23E2.03C IN CHROMOSOME I
8861	21552	34698	0.9	5.1E+00	O16005	SWISSPROT	RHODOPSIN
9725	22376	35577	1.19	5.1E+00	P09182	SWISSPROT	COLICIN N IMMUNITY PROTEIN (MICROCIN N IMMUNITY PROTEIN)
6193	18969	31944	0.72	5.0E+00	BF310443.1	EST_HUMAN	601894910F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4124114 5'
10094	22742		0.59	5.0E+00	BF308561.1	EST_HUMAN	601890420F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4131509 5'
10330	22977	36197	3.07	5.0E+00	AF162445.2	NT	Canis familiaris skeletal muscle chloride channel ClC-1 (CLCN1) mRNA, complete cds
11260	23922	37214	8.95	5.0E+00	Z83860.1	NT	Mycobacterium tuberculosis H37Rv complete genome; segment 103/162
10132	22780		0.71	4.9E+00	U91328.1	NT	Human hereditary haemochromatosis region, histone 2A-like protein gene, hereditary haemochromatosis (HLA-H) gene, RoRet gene, and sodium phosphate transporter (NPT3) gene, complete cds
4039	16784		10.86	4.8E+00	AF185255.1	NT	Eunice australis histone H3 (H3) gene, partial cds
8054	20748	33879	0.47	4.8E+00	BF367909.1	EST_HUMAN	RC3-GN0042-100800-011-c10 GN0042 Homo sapiens cDNA
8439	21131		5.28	4.8E+00	AW750067.1	EST_HUMAN	PM0-BT0547-310100-002-b04 BT0547 Homo sapiens cDNA
283	13090	25731	1.86	4.7E+00	BF240552.1	EST_HUMAN	601875654F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4099716 5'

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
284	13090	25731	1.89	4.7E+00	BF240552.1	EST_HUMAN	601875654F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4099716 5'
3268	16030	28679	2.38	4.7E+00	AL163280.2	NT	Homo sapiens chromosome 21 segment HS21C080
9095	21783	34948	1.18	4.6E+00	BE646437.1	EST_HUMAN	7e68g10.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:3292098 3' similar to TR:O75140 O75140 KIAA0645 PROTEIN. ;contains element PTR5 repetitive element ;
9095	21783	34949	1.18	4.6E+00	BE646437.1	EST_HUMAN	7e68g10.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:3292098 3' similar to TR:O75140 O75140 KIAA0645 PROTEIN. ;contains element PTR5 repetitive element ;
10287	22935		0.61	4.6E+00	AF240786.1	NT	Homo sapiens glutathione S-transferase theta 2 (GSTT2) and glutathione S-transferase theta 1 (GSTT1) genes, complete cds
11054	23724		2.31	4.6E+00	D63999.1	NT	Synechocystis sp. PCC6803 complete genome, 18/27, 2267260-2392728
11605	24204	37526	2.59	4.5E+00	AE001044.1	NT	Archaeoglobus fulgidus section 63 of 172 of the complete genome
11762	24353	37685	1.78	4.5E+00	BF668841.1	EST_HUMAN	602123238F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4280216 5'
3035	15801	28447	0.96	4.4E+00	BF530893.1	EST_HUMAN	602072585F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4215284 5'
3035	15801	28448	0.96	4.4E+00	BF530893.1	EST_HUMAN	602072585F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4215284 5'
6109	18886		1.66	4.4E+00	X13414.1	NT	Murine I gene for MHC class II(Ia) associated invariant chain
6027	18807		0.68	4.3E+00	AF059679.1	NT	Homo sapiens neutrophil collagenase (CLGNA) gene, promoter region and 5'UTR
7338	20019	33097	2.03	4.3E+00	Y13402.1	NT	Plasmodium falciparum R29R+var1 gene, exon 1
7515	20186	33280	0.65	4.3E+00	AE001222.1	NT	Treponema pallidum section 38 of 87 of the complete genome
10769	23453	36696	7.64	4.3E+00	AF240786.1	NT	Homo sapiens glutathione S-transferase theta 2 (GSTT2) and glutathione S-transferase theta 1 (GSTT1) genes, complete cds
5430	18229		3.44	4.2E+00	P16444	SWISSPROT	MICROSOMAL DIPEPTIDASE PRECURSOR (MDP) (DEHYDROPEPTIDASE-I) (RENAL DIPEPTIDASE) (RDP)
5507	18305	31206	0.87	4.2E+00	P51826	SWISSPROT	LAF-4 PROTEIN (LYMPHOID NUCLEAR PROTEIN)
6674	19591	32627	2.62	4.2E+00	P13983	SWISSPROT	EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)
6674	19591	32628	2.62	4.2E+00	P13983	SWISSPROT	EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN)
8859	21550	34697	4.68	4.2E+00	AI809013.1	EST_HUMAN	wf67g03.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2360692 3'
9818	22469	35672	1.06	4.2E+00	P31368	SWISSPROT	NUBBIN PROTEIN (TWAIN PROTEIN) (POU DOMAIN PROTEIN 1) (PDM-1) (DPOU-19) (DOCT1)
10049	22697		0.46	4.2E+00	P40886	SWISSPROT	HEXOSE TRANSPORTER HXT8
5846	25079	31569	0.56	4.1E+00	O09185	SWISSPROT	CELLULAR TUMOR ANTIGEN P53
5846	25079	31570	0.56	4.1E+00	O09185	SWISSPROT	CELLULAR TUMOR ANTIGEN P53
7012	19704	32760	0.84	4.1E+00	BE253668.1	EST_HUMAN	601110727F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3351534 5'
7111	19799	32863	0.65	4.1E+00	BF247939.1	EST_HUMAN	601859030F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4069758 5'
7659	20229	33332	8.73	4.1E+00	O23810	SWISSPROT	YY1 PROTEIN PRECURSOR
7681	20345		0.62	4.1E+00	AB041523.1	NT	Patinopecten yessoensis mRNA for calcineurin A, complete cds
7683	20347	33459	4.32	4.1E+00	P28984	SWISSPROT	GENE 68 PROTEIN

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7683	20347	33460	4.32	4.1E+00	P28964	SWISSPROT	GENE 68 PROTEIN
7817	20512	33638	2.53	4.1E+00	U57503.1	NT	Pan troglodytes novel repetitive solo LTR element in the RNU2 locus
9440	22118	35295	0.67	4.1E+00	P11253	SWISSPROT	50S RIBOSOMAL PROTEIN L4
9571	22224	35409	2.46	4.1E+00	BF692425.1	EST_HUMAN	602247938F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4333209 5'
10205	22853		0.48	4.1E+00	P46414	SWISSPROT	CYCLIN-DEPENDENT KINASE INHIBITOR 1B (CYCLIN-DEPENDENT KINASE INHIBITOR P27) (P27KIP1)
10800	23483		3.06	4.1E+00	P09716	SWISSPROT	HYPOTHETICAL PROTEIN HVLF1
10892	23572		11.69	4.1E+00	BE885880.1	EST_HUMAN	601507510F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3909051 5'
3533	16289		0.95	4.0E+00	P38229	SWISSPROT	GLC7-INTERACTING PROTEIN 1
5372	19500	32524	0.77	4.0E+00	O62653	SWISSPROT	SUCRASE-ISOMALTASE, INTESTINAL [CONTAINS: SUCRASE ; ISOMALTASE]
5372	19500	32525	0.77	4.0E+00	O62653	SWISSPROT	SUCRASE-ISOMALTASE, INTESTINAL [CONTAINS: SUCRASE ; ISOMALTASE]
6838	19500	32524	0.75	4.0E+00	O62653	SWISSPROT	SUCRASE-ISOMALTASE, INTESTINAL [CONTAINS: SUCRASE ; ISOMALTASE]
6838	19500	32525	0.75	4.0E+00	O62653	SWISSPROT	SUCRASE-ISOMALTASE, INTESTINAL [CONTAINS: SUCRASE ; ISOMALTASE]
7089	19778	32843	1.44	4.0E+00	O33010	SWISSPROT	CELL DIVISION PROTEIN FTSY HOMOLOG
8772	21464	34611	0.45	4.0E+00	Q14157	SWISSPROT	HYPOTHETICAL PROTEIN KIAA0144
9843	22494	35695	0.44	4.0E+00	O61309	SWISSPROT	NITRIC-OXIDE SYNTHASE (NOS, TYPE I) (NEURONAL NOS) (N-NOS) (NNOS)
10065	22713	35931	0.63	4.0E+00	AE002132.1	NT	Ureaplasma urealyticum section 33 of 59 of the complete genome
11453	23220	36454	1.53	4.0E+00	P14546	SWISSPROT	CYTOCHROME C OXIDASE POLYPEPTIDE III
11537	24137	37444	2.27	4.0E+00	P07564	SWISSPROT	GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA POLYMERASE (NS5)]
11537	24137	37445	2.27	4.0E+00	P07564	SWISSPROT	GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX PROTEIN (ENVELOPE GLYCOPROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA POLYMERASE (NS5)]
3494	16250	28904	4.61	3.9E+00	X64518.1	NT	N.tabacum chitinase gene 50 for class I chitinase C
4287	17026		8.24	3.9E+00	AF055466.1	NT	Mus musculus seminal vesicle secretory protein 99 (MSVSP99) gene, promoter region
5572	18369	31279	2.91	3.9E+00	BE814357.1	EST_HUMAN	MR0-BN0070-300500-028-h05 BN0070 Homo sapiens cDNA
5572	18369	31280	2.91	3.9E+00	BE814357.1	EST_HUMAN	MR0-BN0070-300500-028-h05 BN0070 Homo sapiens cDNA
6591	19354	32367	0.55	3.9E+00	U91328.1	NT	Human hereditary haemochromatosis region, histone 2A-like protein gene, hereditary haemochromatosis (HLA-H) gene, RoRet gene, and sodium phosphate transporter (NPT3) gene, complete cds
6774	19518	32546	4.62	3.9E+00	P39299	SWISSPROT	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN AIDB-RPSF INTERGENIC REGION

Page 8 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7262	19946	33022	4.3	3.9E+00	M23907.1	NT	Human MHC class II lymphocyte antigen (DPw4-beta-1) gene, exon 2
8216	20910	34046	1.86	3.9E+00	X65865.1	NT	X.laevis mRNA for M4 muscarinic receptor
11365	23176	36403	3.3	3.9E+00	Y18000.1	NT	Homo sapiens NF2 gene
2635	15347		0.9	3.8E+00	AE001562.1	NT	Helicobacter pylori, strain J99 section 123 of 132 of the complete genome
6297	19070	32054	0.96	3.8E+00	Q57830	SWISSPROT	HYPOTHETICAL PROTEIN MJ0385
6673	19590	32626	0.66	3.8E+00	AI493849.1	EST_HUMAN	qz51f07.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2030437 3'
8331	21024	34161	1.1	3.8E+00	D44725.1	EST_HUMAN	HUMSUPY135 Human brain cDNA Homo sapiens cDNA clone 148
9694	22345		0.62	3.8E+00	AJ390961.1	NT	Streptococcus oralis partial xpt gene for xanthine phosphoribosyltransferase, strain NCTC7864
4001	16748	29379	12.29	3.7E+00	AL161539.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 39
7068	19757		1.03	3.7E+00	AL445065.1	NT	Thermoplasma acidophilum complete genome; segment 3/5
8609	21301		0.55	3.7E+00	4503950	NT	Homo sapiens glucokinase (hexokinase 4, maturity onset diabetes of the young 2) (GCK), nuclear gene encoding mitochondrial protein, mRNA
9076	21765	34928	0.7	3.7E+00	U43541.1	NT	Mus musculus laminin beta 2 gene, exons 17-33, and complete cds
11408	24057	37363	2.23	3.7E+00	BF669279.1	EST_HUMAN	602120551F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4277748 5'
11408	24057	37364	2.23	3.7E+00	BF669279.1	EST_HUMAN	602120551F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4277748 5'
579	13359	25986	5.19	3.6E+00	AV761055.1	EST_HUMAN	AV761055 MDS Homo sapiens cDNA clone MDSBUE10 5'
4745	17477		1.06	3.6E+00	AL161472.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 2
5174	17983	30498	0.74	3.6E+00	BF316316.1	EST_HUMAN	601901866F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4131016 5'
8450	21142	34280	0.95	3.6E+00	D12367.1	EST_HUMAN	HUM000TB08 Liver HepG2 cell line. Homo sapiens cDNA clone tb08
8450	21142	34281	0.95	3.6E+00	D12367.1	EST_HUMAN	HUM000TB08 Liver HepG2 cell line. Homo sapiens cDNA clone tb08
8543	21235	34378	3.83	3.6E+00	AE004447.1	NT	Pseudomonas aeruginosa PA01, section 8 of 529 of the complete genome
8543	21235	34379	3.83	3.6E+00	AE004447.1	NT	Pseudomonas aeruginosa PA01, section 8 of 529 of the complete genome
10759	23444		4.07	3.6E+00	M96795.1	NT	Escherichia coli glycerophosphate dehydrogenase (glpD) gene, partial cds; and the translation start site has been verified (glpE), the translation start site has been verified (glpG), and repressor protein (glpR) genes, complete cds
3241	16003	28652	1.1	3.5E+00	AF221538.1	NT	Cryptosporidium felis heat shock protein 70 (HSP70) gene, partial cds
5911	18695		1.17	3.5E+00	L42898.1	NT	Borrelia burgdorferi (strain 25015) outer surface protein (ospC) gene, partial cds
6118	18896	31864	1.18	3.5E+00	R19745.1	EST_HUMAN	yg40c08.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:34940 5'
8383	21076		0.56	3.5E+00	P24557	SWISSPROT	THROMBOXANE-A SYNTHASE (TXA SYNTHASE) (TXS)
8930	21621	34763	1.02	3.5E+00	AA190998.1	EST_HUMAN	zp86b04.s1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone IMAGE:627055 3' similar to contains Alu repetitive element;contains element MSR1 repetitive element ;
8930	21621	34764	1.02	3.5E+00	AA190998.1	EST_HUMAN	zp86b04.s1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone IMAGE:627055 3' similar to contains Alu repetitive element;contains element MSR1 repetitive element ;
9393	22055	35227	0.96	3.5E+00	AL161553.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 53

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10417	23063	36283	0.46	3.5E+00	AJ133723.1	NT	Bos taurus mRNA for Ran-binding protein 2, partial
1501	14247	26933	2.94	3.4E+00	AF254577.1	NT	Brassica napus RPB5d mRNA, complete cds
7261	19945	33021	2.64	3.4E+00	P04052	SWISSPROT	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT
7601	20267	33374	0.69	3.4E+00	P04052	SWISSPROT	DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT
8577	21269		0.7	3.4E+00	U65406.1	NT	Human alternatively spliced potassium channels ROM-K1, ROM-K2, ROM-K3, ROM-K4, ROM-K5, and ROM-K6 (KCNJ1) gene, complete cds
8972	21662	34813	0.67	3.4E+00	AJ229042.1	NT	Homo sapiens 959 kb contig between AML1 and CBR1 on chromosome 21q22, segment 2/3
9010	21700	34850	0.54	3.4E+00	AJ250567.1	NT	Homo sapiens partial TM4SF2 gene for tetraspanin protein, exon 6
10164	22812	36030	2.97	3.4E+00	AF013167.1	NT	Saccharomyces cerevisiae MSS1 gene, complete cds
11519	24119	37429	1.89	3:4E+00	L77570.1	NT	Homo sapiens DiGeorge syndrome critical region, centromeric end
5977	18759	31722	1.57	3.3E+00	Q09669	SWISSPROT	PUTATIVE IRON ALCOHOL DEHYDROGENASE
5977	18759	31723	1.57	3.3E+00	Q09669	SWISSPROT	PUTATIVE IRON ALCOHOL DEHYDROGENASE
7794	20489	33611	0.79	3.3E+00	AF111168.2	NT	Homo sapiens serine palmitoyl transferase, subunit II gene, complete cds; and unknown genes
10361	23008	36223	0.9	3.3E+00	AP001511.1	NT	Bacillus halodurans genomic DNA, section 5/14
10361	23008	36224	0.9	3.3E+00	AP001511.1	NT	Bacillus halodurans genomic DNA, section 5/14
488	13273	25908	1.64	3.2E+00	X96422.1	NT	D.rerio zp-50 POU gene
4004	13273	25908	0.9	3.2E+00	X96422.1	NT	D.rerio zp-50 POU gene
4679	17413	30048	1.08	3.2E+00	4502404	NT	Homo sapiens carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (CEACAM1), mRNA
5481	18280	31176	1.06	3.2E+00	P54924	SWISSPROT	SQUALENE--HOPENE CYCLASE
5481	18280	31177	1.06	3.2E+00	P54924	SWISSPROT	SQUALENE--HOPENE CYCLASE
5515	18313	31214	2.7	3.2E+00	P12783	SWISSPROT	PHOSPHOGLYCERATE KINASE, CYTOSOLIC
5515	18313	31215	2.7	3.2E+00	P12783	SWISSPROT	PHOSPHOGLYCERATE KINASE, CYTOSOLIC
6214	18988	31964	1.78	3.2E+00	P18931	SWISSPROT	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4
6214	18988	31965	1.78	3.2E+00	P18931	SWISSPROT	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4
7505	20176	33270	0.7	3.2E+00	P04275	SWISSPROT	VON WILLEBRAND FACTOR PRECURSOR (VWF)
7672	20336	33448	2.65	3.2E+00	Y13655.1	NT	Chlamydomonas reinhardtii chloroplast DNA for rps9, ycf4, ycf3, rps18 genes
7672	20336	33449	2.65	3.2E+00	Y13655.1	NT	Chlamydomonas reinhardtii chloroplast DNA for rps9, ycf4, ycf3, rps18 genes
8928	21619		4.51	3.2E+00	P13061	SWISSPROT	PERIPLASMIC [NIFE] HYDROGENASE SMALL SUBUNIT (NIFE HYDROGENLYASE SMALL CHAIN)
9430	22108	35283	0.87	3.2E+00	M36383.1	NT	S.cerevisiae threonine deaminase (ILV1) gene, complete cds
10041	22689	35907	2.03	3.2E+00	AB016081.2	NT	Oryzias latipes OlGC8 gene for guanylyl cyclase C, complete cds
11946	24500		2.44	3.2E+00	L33836.1	NT	Sus scrofa choline acetyltransferase gene, promoter region
5785	18576	31505	2.48	3.1E+00	Q10135	SWISSPROT	HYPOTHETICAL 142.5 KD PROTEIN C23E2.02 IN CHROMOSOME I
7287	19970	33047	0.93	3.1E+00	P52178	SWISSPROT	TRIOSE PHOSPHATE/PHOSPHATE TRANSLOCATOR, NON-GREEN PLASTID PRECURSOR (CTPT)

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7627	20293		0.94	3.1E+00	AF303225.1	NT	Bacillus alcalophilus pectate lyase (pelE) gene, complete cds
7986	20681	33807	0.48	3.1E+00	P40985	SWISSPROT	PROBABLE UBIQUITIN--PROTEIN LIGASE HUL4
8500	21192	34333	4.36	3.1E+00	P49894	SWISSPROT	TYPE I IODOTHYRONINE DEIODINASE (TYPE-I 5'DEIODINASE) (DIOI) (TYPE 1 DI) (5DI)
8500	21192	34334	4.36	3.1E+00	P49894	SWISSPROT	TYPE I IODOTHYRONINE DEIODINASE (TYPE-I 5'DEIODINASE) (DIOI) (TYPE 1 DI) (5DI)
9168	21889		3.85	3.1E+00	Q14957	SWISSPROT	GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 3 PRECURSOR (N-METHYL D-ASPARTATE RECEPTOR SUBTYPE 2C) (NR2C) (NMDAR2C)
9796	22447	35652	0.59	3.1E+00	7524759	NT	Chlorella vulgaris chloroplast, complete genome
9888	22538		0.63	3.1E+00	Q10125	SWISSPROT	HYPOTHETICAL 56.3 KD PROTEIN F52C9.5 IN CHROMOSOME III
10234	22882	36095	5.52	3.1E+00	P49365	SWISSPROT	DEOXYHYPUSINE SYNTHASE (DHS)
11440	23207		2.66	3.1E+00	P33515	SWISSPROT	GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA POLYMERASE (NS5)]
11463	24066		3.28	3.1E+00	S56660.1	NT	retinoic acid nuclear receptor isoform beta 2 [mice, embryonal carcinoma cell line, PCC7-MZ1, mRNA, 2971 nt]
2842	15610	28259	1.09	3.0E+00	8923984	NT	Homo sapiens hypothetical protein PRO0889 (PRO0889), mRNA
5254	18060	30689	1.32	3.0E+00	X53096.1	NT	S.aureus genes encoding Sau96I DNA methyltransferase and Sau96I restriction endonuclease
6461	19228	32228	0.83	3.0E+00	X56037.1	NT	Corynebacterium glutamicum thrC gene for threonine synthase (EC 4.2.99.2)
6461	19228	32229	0.83	3.0E+00	X56037.1	NT	Corynebacterium glutamicum thrC gene for threonine synthase (EC 4.2.99.2)
7055	19746		9.09	3.0E+00	P18406	SWISSPROT	CYR61 PROTEIN PRECURSOR (3CH61)
7096	19785		0.6	3.0E+00	Q13201	SWISSPROT	ENDOTHELIAL CELL MULTIMERIN PRECURSOR
8805	21497		1.2	3.0E+00	X67838.1	NT	B.napus DNA for myrosinase
10192	22840	36055	0.62	3.0E+00	Q58605	SWISSPROT	S-ADENOSYLMETHIONINE SYNTHETASE (METHIONINE ADENOSYLTRANSFERASE) (ADOMET SYNTHETASE)
10544	23240	36474	1.57	3.0E+00	Q16181	SWISSPROT	CDC10 PROTEIN HOMOLOG
10931	23611	36860	6.44	3.0E+00	P51842	SWISSPROT	RETINAL GUANYLYL CYCLASE 2 PRECURSOR (GUANYLATE CYCLASE 2F, RETINAL) (RETGC-2) (ROD OUTER SEGMENT MEMBRANE GUANYLATE CYCLASE 2) (ROS-GC2) (GUANYLATE CYCLASE F) (GC-F)
10931	23611	36861	6.44	3.0E+00	P51842	SWISSPROT	RETINAL GUANYLYL CYCLASE 2 PRECURSOR (GUANYLATE CYCLASE 2F, RETINAL) (RETGC-2) (ROD OUTER SEGMENT MEMBRANE GUANYLATE CYCLASE 2) (ROS-GC2) (GUANYLATE CYCLASE F) (GC-F)
11578	24177	37492	2.72	3.0E+00	P34194	SWISSPROT	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4
2004	14740	27464	2.28	2.9E+00	AE002225.2	NT	Chlamydophila pneumoniae AR39, section 53 of 94 of the complete genome
6809	19470	32493	1.74	2.9E+00	Z36879.1	NT	F.pringlei gdcsPA gene for P-protein of the glycine cleavage system
7110	19798	32881	5.21	2.9E+00	O14514	SWISSPROT	BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 1 PRECURSOR

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7110	19798	32862	5.21	2.9E+00	O14514	SWISSPROT	BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 1 PRECURSOR
7356	20037	33115	6.84	2.9E+00	P46589	SWISSPROT	ADHERENCE FACTOR (ADHESION AND AGGREGATION MEDIATING SURFACE ANTIGEN)
7767	20463	33587	0.67	2.9E+00	P05844	SWISSPROT	STRUCTURAL POLYPROTEIN [CONTAINS: MAJOR STRUCTURAL PROTEIN VP2; NONSTRUCTURAL PROTEIN VP4; MINOR STRUCTURAL PROTEIN VP3]
7767	20463	33588	0.67	2.9E+00	P05844	SWISSPROT	STRUCTURAL POLYPROTEIN [CONTAINS: MAJOR STRUCTURAL PROTEIN VP2; NONSTRUCTURAL PROTEIN VP4; MINOR STRUCTURAL PROTEIN VP3]
7996	20691	33819	1.03	2.9E+00	BF344171.1	EST_HUMAN	602017413F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4163059 5'
1440	14187	26872	4.4	2.8E+00	AF186398.1	NT	Buxus harlandii maturase K (matK) gene, partial cds; chloroplast gene for chloroplast product
1629	14375		2.74	2.8E+00	AL161552.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 52
7207	19892	32968	5.72	2.8E+00	8393724	NT	Mus musculus endomucin (LOC53423), mRNA
9513	22166		0.54	2.8E+00	BE565182.1	EST_HUMAN	601342758F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3684807 5'
10588	19892	32968	1.32	2.8E+00	8393724	NT	Mus musculus endomucin (LOC53423), mRNA
224	13036	25672	13.51	2.7E+00	6679306	NT	Mus musculus per-hexamer repeat gene 3 (Phxr3), mRNA
224	13036	25673	13.51	2.7E+00	6679306	NT	Mus musculus per-hexamer repeat gene 3 (Phxr3), mRNA
5484	18263	31154	1.17	2.7E+00	L14005.1	NT	Homo sapiens apoA polymorphism Kringle IV gene, exons 1 and 2
8045	20739		0.6	2.7E+00	U15947.1	NT	Ipomoea purpurea chalcone synthase (CHSB) gene including complete 5'UTR and complete cds
8867	21558		1.83	2.7E+00	AL116459.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
9332	20403	33519	0.73	2.7E+00	AW088191.1	EST_HUMAN	xc88e12.x1 NCI_CGAP_Brn35 Homo sapiens cDNA clone IMAGE:2591374 3' similar to gb:M17733 THYMOSIN BETA-4 (HUMAN);
10397	23043		1.75	2.7E+00	BE063527.1	EST_HUMAN	CM0-BT0281-031199-087-h04 BT0281 Homo sapiens cDNA
4626	17361	29994	5.15	2.6E+00	AF068749.1	NT	Mus musculus sphingosine kinase (SPHK1b) mRNA, complete cds
5460	18259	31149	1.68	2.6E+00	6755601	NT	Mus musculus SRY-box containing gene 13 (Sox13), mRNA
5460	18259	31150	1.68	2.6E+00	6755601	NT	Mus musculus SRY-box containing gene 13 (Sox13), mRNA
5736	18528		0.59	2.6E+00	Y17062.1	NT	Mycobacterium fortuitum furA II gene
7454	25424		0.82	2.6E+00	AJ224639.1	NT	Homo sapiens Surf-5 and Surf-6 genes
7600	20266		6.04	2.6E+00	AF235502.1	NT	Mus musculus SH2-containing inositol 5-phosphatase (Ship) gene, exons 16 through 27, and complete cds
7958	20653	33776	1.13	2.6E+00	AJ132180.1	NT	faba bean necrotic yellows virus C2-Eg gene, isolate Egyptian EV1-93
7958	20653	33777	1.13	2.6E+00	AJ132180.1	NT	faba bean necrotic yellows virus C2-Eg gene, isolate Egyptian EV1-93
9557	22210	35395	2.83	2.6E+00	AL161540.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 40
10253	22901		1.67	2.6E+00	9055193	NT	Mus musculus cleavage and polyadenylation specificity factor 3 (Cpsf3), mRNA
10953	23630	36878	1.32	2.6E+00	AF143675.1	NT	Hantavirus Z10 segment M G1/G2 glycoprotein (Z10) gene, complete cds
12560	25304		3.17	2.6E+00	11419220	NT	Homo sapiens ATP-binding cassette, sub-family B (MDR/TAP), member 4 (ABCB4), mRNA
1448	14195	26878	3.73	2.5E+00	AJ271844.1	NT	Aspergillus nidulans recQ gene for DNA helicase, exons 1-4

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1448	14195	26879	3.73	2.5E+00	AJ271844.1	NT	Aspergillus nidulans recQ gene for DNA helicase, exons 1-4
5723	18515	31434	2.22	2.5E+00	P13485	SWISSPROT	TEICHOIC ACID BIOSYNTHESIS PROTEIN F
5723	18515	31435	2.22	2.5E+00	P13485	SWISSPROT	TEICHOIC ACID BIOSYNTHESIS PROTEIN F
6367	18515	31434	1.63	2.5E+00	P13485	SWISSPROT	TEICHOIC ACID BIOSYNTHESIS PROTEIN F
6367	18515	31435	1.63	2.5E+00	P13485	SWISSPROT	TEICHOIC ACID BIOSYNTHESIS PROTEIN F
6630	19392	32408	0.64	2.5E+00	D30052.1	NT	Vibrio cholerae ctxA gene and ctxB gene for cholera toxins, complete cds
7659	20323	33431	0.99	2.5E+00	AW949158.1	EST_HUMAN	QV4-FT0005-110500-205-g07 FT0005 Homo sapiens cDNA
7700	20363	33477	0.58	2.5E+00	4502902	NT	Homo sapiens clathrin, heavy polypeptide-like 1 (CLTCL1) mRNA
9001	21691	34841	1.53	2.5E+00	D50307.1	NT	Rice DNA for aldolase C-1, complete cds
9752	22403	35608	0.67	2.5E+00	BE297758.1	EST_HUMAN	601175779F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3531090 5'
11528	24128		1.34	2.5E+00	P40170	SWISSPROT	DNAJ PROTEIN
11943	24498		3.08	2.5E+00	AF289665.1	NT	Mus musculus EIF4H gene, partial cds; LIMK1 gene, complete cds; and ELN gene, partial cds
3012	15778	28428	1.13	2.4E+00	M24282.1	NT	Chicken alpha-3 collagen type VI mRNA, 3' end
4849	17579	30203	6.09	2.4E+00	4503352	NT	Homo sapiens double C2-like domains, alpha (DOC2A) mRNA
5920	18705	31657	4.16	2.4E+00	P02843	SWISSPROT	VITELLOGENIN I PRECURSOR (YOLK PROTEIN 1)
7280	19964	33040	0.78	2.4E+00	BF667502.1	EST_HUMAN	602120856F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4278012 5'
7280	19964	33041	0.78	2.4E+00	BF667502.1	EST_HUMAN	602120856F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4278012 5'
8039	20734	33865	2.4	2.4E+00	P26842	SWISSPROT	CD27L RECEPTOR PRECURSOR (T-CELL ACTIVATION ANTIGEN CD27) (T14)
8039	20734	33866	2.4	2.4E+00	P26842	SWISSPROT	CD27L RECEPTOR PRECURSOR (T-CELL ACTIVATION ANTIGEN CD27) (T14)
8110	20804		2.63	2.4E+00	AE001486.1	NT	Helicobacter pylori, strain J99 section 47 of 132 of the complete genome
8549	21241		1.61	2.4E+00	AW875126.1	EST_HUMAN	RC2-PT0004-031299-011-d05 PT0004 Homo sapiens cDNA
8727	21419	34563	7.36	2.4E+00	P24091	SWISSPROT	ENDOCHITINASE B PRECURSOR (CHN-B)
9938	22586	35788	2.56	2.4E+00	P13673	SWISSPROT	SKIN GRANULE PROTEIN PRECURSOR
9938	22586	35789	2.56	2.4E+00	P13673	SWISSPROT	SKIN GRANULE PROTEIN PRECURSOR
10007	22655	35868	1.86	2.4E+00	X92511.1	NT	H.sapiens CTGF gene and promoter region
10143	22791		6.65	2.4E+00	P09099	SWISSPROT	XYLULOSE KINASE (XYLULOKINASE)
10220	22868	36079	1.62	2.4E+00	BE326702.1	EST_HUMAN	hr63f06.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3133187 3'
10220	22868	36080	1.62	2.4E+00	BE326702.1	EST_HUMAN	hr63f06.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3133187 3'
10490	23136	36364	0.87	2.4E+00	Q51481	SWISSPROT	DENITRIFICATION REGULATORY PROTEIN NIRQ
11331	24022	37327	2.16	2.4E+00	AF158652.2	NT	Fragaria x ananassa cytosolic ascorbate peroxidase (ApxSC) gene, ApxSC-c allele, complete cds
1231	13980	26650	13.6	2.3E+00	Z46724.1	NT	G.domesticus artificial single chain antibody gene (L3)
4102	16845		1.35	2.3E+00	AJ401081.1	NT	Bos taurus partial cytb gene for cytochrome b

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5744	18536		0.95	2.3E+00	N88245.1	EST_HUMAN	J7340F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA clone J7340 5' similar to PROLYLCARBOXYPEPTIDASE
7354	20035	33113	2.47	2.3E+00	6978554	NT	Rattus norvegicus ATPase, Ca++ transporting, ubiquitous (Atp2a3), mRNA
7495	25425		3.07	2.3E+00	P07199	SWISSPROT	MAJOR CENTROMERE AUTOANTIGEN B (CENTROMERE PROTEIN B) (CENP-B)
7679	20343	33455	1.01	2.3E+00	X60265.1	NT	M.mazei dnaK and dnaJ genes homologues coding for DnaK and DnaJ
9008	21698	34849	0.54	2.3E+00	5835317	NT	Polypterus ornatipinnis mitochondrion, complete genome
9068	21757	34919	1.8	2.3E+00	Q11127	SWISSPROT	ALPHA-(1,3)-FUCOSYLTRANSFERASE (GALACTOSIDE 3-L-FUCOSYLTRANSFERASE) (FUCOSYLTRANSFERASE 4) (FUCT-IV)
10704	23395	36632	3.83	2.3E+00	Q07076	SWISSPROT	ANNEXIN VII (SYNEXIN)
11782	24373	37703	3.03	2.3E+00	BF541987.1	EST_HUMAN	602069121F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4068173 5'
11782	24373	37704	3.03	2.3E+00	BF541987.1	EST_HUMAN	602069121F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4068173 5'
12157	24642	31099	6.84	2.3E+00	BE895237.1	EST_HUMAN	601433673F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3918643 5'
3998	16746	29378	0.95	2.2E+00	AF020528.1	NT	Magnaporthe grisea Class IV chitin synthase (chs4) gene, complete cds
4278	17017	29644	5.01	2.2E+00	D67071.1	NT	Rat gene for regucalcin, exon1 (non-coding exon)
4278	17017	29645	5.01	2.2E+00	D67071.1	NT	Rat gene for regucalcin, exon1 (non-coding exon)
5258	18064	30692	12.73	2.2E+00	O88307	SWISSPROT	SORTILIN-RELATED RECEPTOR PRECURSOR (SORTING PROTEIN-RELATED RECEPTOR CONTAINING LDLR CLASS A REPEATS) (MSORLA) (SORLA-1) (LOW-DENSITY LIPOPROTEIN RECEPTOR RELATIVE WITH 11 LIGAND-BINDING REPEATS) (LDLR RELATIVE WITH 11 LIGAND-BINDING REPEATS) (LR11) (>
5258	18064	30693	12.73	2.2E+00	O88307	SWISSPROT	SORTILIN-RELATED RECEPTOR PRECURSOR (SORTING PROTEIN-RELATED RECEPTOR CONTAINING LDLR CLASS A REPEATS) (MSORLA) (SORLA-1) (LOW-DENSITY LIPOPROTEIN RECEPTOR RELATIVE WITH 11 LIGAND-BINDING REPEATS) (LDLR RELATIVE WITH 11 LIGAND-BINDING REPEATS) (LR11) (>
5763	18554	31478	1.03	2.2E+00	BE927220.1	EST_HUMAN	RC3-CT0254-300800-022-e08 CT0254 Homo sapiens cDNA
5763	18554	31479	1.03	2.2E+00	BE927220.1	EST_HUMAN	RC3-CT0254-300800-022-e06 CT0254 Homo sapiens cDNA
5971	18753	31714	9.84	2.2E+00	BE250383.1	EST_HUMAN	600943401T1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:2959777 3'
6261	19035	32010	3.87	2.2E+00	Q00335	SWISSPROT	MINOR VIRION STRUCTURAL PROTEIN MU-2
6502	19267	32269	3.16	2.2E+00	P51459	SWISSPROT	INSULIN-LIKE GROWTH FACTOR II PRECURSOR (IGF-II) (SOMATOMEDIN A)
6861	17938		3.94	2.2E+00	AA594674.1	EST_HUMAN	nl95b02.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1058379 3'
7217	19902	32975	0.9	2.2E+00	AA137027.1	EST_HUMAN	zn97f04.r1 Stratagene fetal retina 937202 Homo sapiens cDNA clone IMAGE:566143 5'
7507	20178	33272	19.2	2.2E+00	AA449012.1	EST_HUMAN	zx05g10.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:785634 5'
7589	20257	33365	0.72	2.2E+00	P54918	SWISSPROT	ALANINE RACEMASE
8001	20696	33823	0.58	2.2E+00	BE301560.1	EST_HUMAN	bb17h12.x1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:2963207 3' similar to gb:D45836 Mouse mRNA for nuclear pore-targeting-complex component of (MOUSE);

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8001	20696	33824	0.58	2.2E+00	BE301560.1	EST_HUMAN	bb17h12.x1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:2963207 3' similar to gb:D45836 Mouse mRNA for nuclear pore-targeting-complex component of (MOUSE);
9241	21920		11.02	2.2E+00	BE741678.1	EST_HUMAN	601594733F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3948561 5'
9468	25124		2.28	2.2E+00	Q04706	SWISSPROT	TRANSPOSON TY1 PROTEIN A
9953	22601	35804	1.1	2.2E+00	AI290373.1	EST_HUMAN	qm69b03.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:1893965 3' similar to gb:Y00433 GLUTATHIONE PEROXIDASE (HUMAN);
9953	22601	35805	1.1	2.2E+00	AI290373.1	EST_HUMAN	qm69b03.x1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:1893965 3' similar to gb:Y00433 GLUTATHIONE PEROXIDASE (HUMAN);
9996	22644	35858	2.68	2.2E+00	BF246782.1	EST_HUMAN	601855591F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4075391 5'
10353	23000	36217	3.11	2.2E+00	AF183416.1	NT	Homo sapiens ovarian granulosa cell 13.0 kDa protein hGR74 homolog mRNA, complete cds
11418	23185	36415	3.47	2.2E+00	P07911	SWISSPROT	UROMODULIN PRECURSOR (TAMM-HORSFALL URINARY GLYCOPROTEIN) (THP)
11616	24214	37539	5.89	2.2E+00	P10407	SWISSPROT	EARLY E1A 28 KD PROTEIN
556	15545	25967	8.3	2.1E+00	AF132612.2	NT	Mus musculus pre-T cell receptor alpha gene, enhancer region and upstream region
3575	16330		1.08	2.1E+00	AW449366.1	EST_HUMAN	UI-H-BI3-akl-e-08-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone IMAGE:2734550 3'
6041	18821		0.89	2.1E+00	P75357	SWISSPROT	HYPOTHETICAL PROTEIN MG302 HOMOLOG
6710	19625	32689	3.95	2.1E+00	O70159	SWISSPROT	ALPHA-2-HS-GLYCOPROTEIN PRECURSOR (FETUIN-A)
6946	19428	32443	5.72	2.1E+00	N29575.1	EST_HUMAN	yy08a10.s1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:270618 3' similar to gb:M55654 TRANSCRIPTION INITIATION FACTOR TFIID (HUMAN);
8395	21088		1.97	2.1E+00	AU123630.1	EST_HUMAN	AU123630 NT2RM2 Homo sapiens cDNA clone NT2RM2000671 5'
1174	13927	26591	1.44	2.0E+00	AF180527.1	NT	Homo sapiens p22Dokdel (DOKDEL) mRNA, complete cds
1174	13927	26592	1.44	2.0E+00	AF180527.1	NT	Homo sapiens p22Dokdel (DOKDEL) mRNA, complete cds
1312	14060	26735	0.97	2.0E+00	AF204927.1	NT	Oryctolagus cuniculus Na+,K+-ATPase beta 1 subunit mRNA, complete cds
1569	14316		2.61	2.0E+00	P25582	SWISSPROT	PUTATIVE RRNA METHYLTRANSFERASE SPB1
2145	14875	27609	5.98	2.0E+00	Z78279.1	NT	R.norvegicus mRNA for collagen alpha1 type I
2145	14875	27610	5.98	2.0E+00	Z78279.1	NT	R.norvegicus mRNA for collagen alpha1 type I
4080	16824	29450	2.2	2.0E+00	AW684498.1	EST_HUMAN	hi13c05.x1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2972168 3' similar to gb:X01677 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, LIVER (HUMAN);
4080	16824	29451	2.2	2.0E+00	AW684498.1	EST_HUMAN	hi13c05.x1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2972168 3' similar to gb:X01677 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, LIVER (HUMAN);
7449	20125		0.92	2.0E+00	P07566	SWISSPROT	STRUCTURAL POLYPROTEIN [CONTAINS: NUCLEOCAPSID PROTEIN C; MEMBRANE GLYCOPROTEINS E1 AND E2]
7923	20618	33745	3.17	2.0E+00	AB008676.1	NT	Escherichia coli 0157 DNA, map position at 46 min., complete cds
7923	20618	33746	3.17	2.0E+00	AB008676.1	NT	Escherichia coli 0157 DNA, map position at 46 min., complete cds
7923	20618	33747	3.17	2.0E+00	AB008676.1	NT	Escherichia coli 0157 DNA, map position at 46 min., complete cds

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8820	21512	34656	3.15	2.0E+00	F31500.1	EST_HUMAN	HSPD22703 HM3 Homo sapiens cDNA clone s4000117B08
12481	25265	30720	7.27	2.0E+00	5834843	NT	Gallus gallus mitochondrion, complete genome
5511	18309	31209	4.77	1.9E+00	6754389	NT	Mus musculus inositol 1,4,5-triphosphate receptor 1 (Itpr1), mRNA
5511	18309	31210	4.77	1.9E+00	6754389	NT	Mus musculus inositol 1,4,5-triphosphate receptor 1 (Itpr1), mRNA
6009	18790	31753	1.32	1.9E+00	BE969695.1	EST_HUMAN	601679636F1 NIH_MGC_78 Homo sapiens cDNA clone IMAGE:3949881 5'
6556	19321		0.75	1.9E+00	AW845689.1	EST_HUMAN	MR0-CT0063-071099-002-g02 CT0063 Homo sapiens cDNA
6650	19412		2.46	1.9E+00	Q63627	SWISSPROT	CTD-BINDING SR-LIKE PROTEIN RA4
8358	21051	34190	2.18	1.9E+00	P02467	SWISSPROT	COLLAGEN ALPHA 2(I) CHAIN PRECURSOR
8358	21051	34191	2.18	1.9E+00	P02467	SWISSPROT	COLLAGEN ALPHA 2(I) CHAIN PRECURSOR
8557	21249		2.94	1.9E+00	BF360206.1	EST_HUMAN	CM3-MT0114-010900-323-h12 MT0114 Homo sapiens cDNA
8792	21484		1.33	1.9E+00	O51781	SWISSPROT	ARGININE DEIMINASE (ADI) (ARGININE DIHYDROLASE) (AD)
9530	22183	35367	0.59	1.9E+00	AA669125.1	EST_HUMAN	ab94a04.s1 Stratagene lung (#937210) Homo sapiens cDNA clone IMAGE:854574 3' similar to contains Alu repetitive element;contains element L1 L1 repetitive element ;
10462	23108	36339	0.62	1.9E+00	AF248269.1	NT	Homo sapiens gag-pro-pol precursor protein gene, partial cds
3089	15854	28496	1.3	1.8E+00	P21004	SWISSPROT	PROTEIN B8 PRECURSOR
3118	15883	28522	1.57	1.8E+00	U04356.1	NT	Synechococcus sp. PCC7942 copper transporting P-ATPase (ctaA) and ATP synthase epsilon subunit (atpE) genes, complete cds
3118	15883	28523	1.57	1.8E+00	U04356.1	NT	Synechococcus sp. PCC7942 copper transporting P-ATPase (ctaA) and ATP synthase epsilon subunit (atpE) genes, complete cds
5777	18568		1.91	1.8E+00	P18502	SWISSPROT	HEDGEHOG RECEPTOR (PATCHED PROTEIN)
6013	18794	31757	1.32	1.8E+00	BF311999.1	EST_HUMAN	601897854F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4127364 5'
6305	19077		1.12	1.8E+00	BF683327.1	EST_HUMAN	602139470F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4298272 5'
6641	19403	32418	1.64	1.8E+00	BF305652.1	EST_HUMAN	601893489F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4139038 5'
6958	19440	32455	1.79	1.8E+00	P21249	SWISSPROT	MAJOR ANTIGEN
8016	20711	33841	0.93	1.8E+00	P11369	SWISSPROT	RETROVIRUS-RELATED POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE ; ENDONUCLEASE]
8016	20711	33842	0.93	1.8E+00	P11369	SWISSPROT	RETROVIRUS-RELATED POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE ; ENDONUCLEASE]
8368	21061	34201	0.44	1.8E+00	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
8368	21061	34202	0.44	1.8E+00	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
8368	21061	34203	0.44	1.8E+00	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
8753	21445	34593	1.98	1.8E+00	O43281	SWISSPROT	EMBRYONAL FYN-ASSOCIATED SUBSTRATE (HEFS)
9073	21762	34924	0.77	1.8E+00	R31042.1	EST_HUMAN	yh72c08.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:135278 5'
9161	21831	34994	0.76	1.8E+00	AW880004.1	EST_HUMAN	QV0-OT0030-070300-148-a03 OT0030 Homo sapiens cDNA

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9749	22400	35605	0.75	1.8E+00	P27050	SWISSPROT	CHITINASE D PRECURSOR
10183	22831		3.2	1.8E+00	AF111849.1	NT	Homo sapiens PRO0530 mRNA, complete cds
10452	23098		0.63	1.8E+00	P44325	SWISSPROT	CYTIDINE DEAMINASE (CYTIDINE AMINOHYDROLASE) (CDA)
12276	25236		5.29	1.8E+00	AF314254.1	NT	Chlamydomonas reinhardtii alternative oxidase 1 (AOX1) gene, nuclear gene encoding mitochondrial protein
12359	24763		3.9	1.8E+00	9506404	NT	Rattus norvegicus Actin-related protein complex 1b (Arpc1b), mRNA
1086	13844	26502	2.21	1.7E+00	Q60114	SWISSPROT	LEVANSUCRASE (BETA-D-FRUCTOFURANOSYL TRANSFERASE) (SUCROSE 6-FRUCTOSYL TRANSFERASE)
2269	14995	27734	2.29	1.7E+00	AL163280.2	NT	Homo sapiens chromosome 21 segment HS21C080
2372	15094	27833	2.56	1.7E+00	AI141067.1	EST_HUMAN	oz43h05.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1678137 3'
4426	17162	29792	0.81	1.7E+00	Q60114	SWISSPROT	LEVANSUCRASE (BETA-D-FRUCTOFURANOSYL TRANSFERASE) (SUCROSE 6-FRUCTOSYL TRANSFERASE)
5525	18323	31223	1.77	1.7E+00	BE063546.1	EST_HUMAN	CM0-BT0282-171299-127-e05 BT0282 Homo sapiens cDNA
5525	18323	31224	1.77	1.7E+00	BE063546.1	EST_HUMAN	CM0-BT0282-171299-127-e05 BT0282 Homo sapiens cDNA
5927	18711	31666	3.28	1.7E+00	Q9TTR8	SWISSPROT	COUP TRANSCRIPTION FACTOR 1 (COUP-TF1) (COUP-TF I)
7118	19806	32871	1.11	1.7E+00	Q03703	SWISSPROT	HYPOTHETICAL 38.0 KD PROTEIN IN CAT2-AMD1 INTERGENIC REGION
7118	19806	32872	1.11	1.7E+00	Q03703	SWISSPROT	HYPOTHETICAL 38.0 KD PROTEIN IN CAT2-AMD1 INTERGENIC REGION
7753	20449	33573	0.91	1.7E+00	AF021335.1	NT	Mus musculus T cell receptor gamma locus, TCR gamma 2 and gamma 4 gene clusters
7932	20627	33755	1.13	1.7E+00	6755715	NT	Mus musculus T-cell acute lymphocytic leukemia 1 (Tal1), mRNA
7961	20656	33781	0.59	1.7E+00	BF530630.1	EST_HUMAN	602071917F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4214669 5'
8440	21132	34268	0.6	1.7E+00	AF245513.1	NT	Hippoglossus hippoglossus interferon inducible Mx protein (Mx) mRNA, complete cds
8525	21217		2.3	1.7E+00	BF308000.1	EST_HUMAN	601894255F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4140084 5'
8605	21297	34440	0.59	1.7E+00	X69063.1	NT	M.musculus Ank-1 mRNA for erythroid ankyrin
8605	21297	34441	0.59	1.7E+00	X69063.1	NT	M.musculus Ank-1 mRNA for erythroid ankyrin
9047	25123	34892	2.18	1.7E+00	O60479	SWISSPROT	HOMEOBOX PROTEIN DLX-3
9047	25123	34893	2.18	1.7E+00	O60479	SWISSPROT	HOMEOBOX PROTEIN DLX-3
9506	22159		1.15	1.7E+00	AF161380.1	NT	Homo sapiens HSPC262 mRNA, partial cds
10071	22719		0.48	1.7E+00	AW953681.1	EST_HUMAN	EST365751 MAGE resequences, MAGC Homo sapiens cDNA
11596	24195	37514	2.57	1.7E+00	W22424.1	EST_HUMAN	67B7 Human retina cDNA Tsp509I-cleaved sublibrary Homo sapiens cDNA not directional
12231	24684	31074	1.9	1.7E+00	AI678443.1	EST_HUMAN	tu82d07.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2257549 3' similar to contains MSR1.t1 MSR1 repetitive element ;
12717	24990	30970	1.84	1.7E+00	AI198573.1	EST_HUMAN	qf50b01.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1753417 3' similar to contains L1.t1 L1 repetitive element ;
2027	14762	27491	18.51	1.6E+00	AF199339.1	NT	Homo sapiens lens epithelium-derived growth factor gene, alternatively spliced, complete cds

Page 17 of 536

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2036	14771	27500	3.75	1.6E+00	AF077374.1	NT	Homo sapiens small proline-rich protein (SPRR3) gene, exons 1, 2, and 3 and complete cds
2042	14776	27505	1.54	1.6E+00	Y11344.1	NT	Mus musculus ST6GalNAcIII gene, exon 2
2282	15007		1.24	1.6E+00	X98373.1	NT	B.napus gene encoding endo-polygalacturonase
2961	15727	28377	1.61	1.6E+00	W58426.1	EST_HUMAN	zd25f01.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:341689 5' similar to gb:D29805 N-ACETYLLACTOSAMINE SYNTHASE (HUMAN);
4011	16757		5.66	1.6E+00	BF570077.1	EST_HUMAN	602186095T1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4310591 3'
4319	17058	29682	1.9	1.6E+00	AF155827.1	NT	Homo sapiens proliferation-associated SNF2-like protein (SMARCA6) mRNA, complete cds
4319	17058	29683	1.9	1.6E+00	AF155827.1	NT	Homo sapiens proliferation-associated SNF2-like protein (SMARCA6) mRNA, complete cds
4942	17669	30277	0.84	1.6E+00	AF075394.1	NT	Uroteuthis chinensis cytochrome c oxidase subunit I (COI) gene, mitochondrial gene encoding mitochondrial protein, partial cds
4942	17669	30278	0.84	1.6E+00	AF075394.1	NT	Uroteuthis chinensis cytochrome c oxidase subunit I (COI) gene, mitochondrial gene encoding mitochondrial protein, partial cds
5024	17745	30356	2.86	1.6E+00	Y11344.1	NT	Mus musculus ST6GalNAcIII gene, exon 2
5024	17745	30357	2.86	1.6E+00	Y11344.1	NT	Mus musculus ST6GalNAcIII gene, exon 2
5737	18529	31450	2.16	1.6E+00	L04808.1	NT	Brachydanio rerio MHC class II DA-beta-2*01 gene, 3' end
5823	18612	31543	0.79	1.6E+00	AF005631.1	NT	Homo sapiens transglutaminase type I (TgaseI) gene, promoter region
6378	19147	32146	0.69	1.6E+00	BF380703.1	EST_HUMAN	IL2-UT0073-060900-145-E02 UT0073 Homo sapiens cDNA
6610	19373	32387	1.06	1.6E+00	AW294881.1	EST_HUMAN	UI-H-BI2-ahr-b-04-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone IMAGE:2727511 3'
7145	19832	32901	2.73	1.6E+00	BE697267.1	EST_HUMAN	RC0-CT0416-200700-032-c10 CT0416 Homo sapiens cDNA
7929	20624		1.19	1.6E+00	Q46378	SWISSPROT	VIRULENCE FACTOR MVIN HOMOLOG
8277	20971	34112	3.28	1.6E+00	AJ297131.1	NT	Mus musculus SIL, MAP_17, CYP_a, SCL & CYP_b genes
8798	21490	34636	0.83	1.6E+00	11437222	NT	Homo sapiens hypothetical protein PRO0971 (PRO0971), mRNA
8798	21490	34637	0.83	1.6E+00	11437222	NT	Homo sapiens hypothetical protein PRO0971 (PRO0971), mRNA
8970	21660	34810	0.47	1.6E+00	BE388331.1	EST_HUMAN	601283925F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3605647 5'
9360	25121	33549	1.94	1.6E+00	X52046.1	NT	M.musculus COL3A1 gene for collagen alpha-I
9360	25121	33550	1.94	1.6E+00	X52046.1	NT	M.musculus COL3A1 gene for collagen alpha-I
9487	22140		0.56	1.6E+00	AF043468.1	NT	Thermoanaerobacter ethanolicus D-xylose-binding protein (xylF) gene, complete cds
9634	22288	35480	1.32	1.6E+00	T41290.1	EST_HUMAN	ph6b6_19/1TV Outward Alu-primed hncDNA library Homo sapiens cDNA clone ph6b6_19/1TV
10047	22695	35911	0.5	1.6E+00	AF121361.1	NT	Drosophila melanogaster signal transducting adaptor protein (STAM), serine threonine kinase lal (IAL), and zinc finger protein (DNZ1) genes, complete cds
10085	22733	35947	1.15	1.6E+00	AW835644.1	EST_HUMAN	QV4-LT0016-090200-100-d07 LT0016 Homo sapiens cDNA
10085	22733	35948	1.15	1.6E+00	AW835644.1	EST_HUMAN	QV4-LT0016-090200-100-d07 LT0016 Homo sapiens cDNA
10242	22890	36102	0.47	1.6E+00	AF037352.1	NT	Mus musculus T cell receptor gamma locus, TCR gamma 1 and gamma 3 gene clusters
10491	23137	36365	0.45	1.6E+00	AF162084.1	NT	Glugea plecoglossi beta-tubulin 2 (btub2) gene, partial cds

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10670	23361	36602	1.95	1.6E+00	P54817	SWISSPROT	CAPSID PROTEIN P40 [CONTAINS: ASSEMBLIN (PROTEASE) ; CAPSID ASSEMBLY PROTEIN]
10728	23416	36657	1.27	1.6E+00	AA216387.1	EST_HUMAN	nc16b02.s1 NCI_CGAP_Pr1 Homo sapiens cDNA clone IMAGE:1008267 similar to contains element MER4 repetitive element ;
10747	18612	31543	5.27	1.6E+00	AF005631.1	NT	Homo sapiens transglutaminase type I (Tgase1) gene, promoter region
11705	24300	37626	3.46	1.6E+00	AF104313.1	NT	Homo sapiens unknown mRNA
31	12859	25476	5.31	1.5E+00	U53449.1	NT	Rattus norvegicus jun dimerization protein 2 (jdp-2) mRNA, complete cds
225	13037	25674	2.2	1.5E+00	AE002201.2	NT	Chlamydophila pneumoniae AR39, section 32 of 94 of the complete genome
606	13384		2.03	1.5E+00	6752961	NT	Mus musculus a disintegrin and metalloproteinase domain (ADAM) 15 (metargidin) (Adam15), mRNA
2410	15131	27867	1.95	1.5E+00	AJ131402.1	NT	Potato virus A RNA complete genome, isolate U
2519	15235	27976	2	1.5E+00	6678350	NT	Mus musculus T-cell lymphoma invasion and metastasis 1 (Tiam1), mRNA
3135	15131	27867	1.85	1.5E+00	AJ131402.1	NT	Potato virus A RNA complete genome, isolate U
3368	16127	28785	0.72	1.5E+00	AE001945.1	NT	Deinococcus radiodurans R1 section 82 of 229 of the complete chromosome 1
5642	18437	31350	0.83	1.5E+00	AI655301.1	EST_HUMAN	tt12f10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2240587 3' similar to TR:O00237 O00237 HKF-1. ;
5642	18437	31351	0.83	1.5E+00	AI655301.1	EST_HUMAN	tt12f10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2240587 3' similar to TR:O00237 O00237 HKF-1. ;
6312	19083	32068	3.02	1.5E+00	R17879.1	EST_HUMAN	yg10e02.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:31693 5'
7028	19720		1.37	1.5E+00	BE785356.1	EST_HUMAN	601478745F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3881555 5'
7060	19751	32814	23.98	1.5E+00	P47179	SWISSPROT	HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DAL5 INTERGENIC REGION PRECURSOR
7060	19751	32815	23.98	1.5E+00	P47179	SWISSPROT	HYPOTHETICAL 118.4 KD PROTEIN IN BAT2-DAL5 INTERGENIC REGION PRECURSOR
7245	19930	33006	0.61	1.5E+00	AA889259.1	EST_HUMAN	ak26f10.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1407115 3'
7493	20165	33257	0.78	1.5E+00	AI003254.1	EST_HUMAN	an07b11.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone IMAGE:1684893 3' similar to gb:S95936 SEROTRANSFERRIN PRECURSOR (HUMAN);
7727	20390		0.64	1.5E+00	AB039887.1	NT	Homo sapiens WDR4 gene for WD repeat protein, complete cds
8021	20716	33848	0.89	1.5E+00	BE887446.1	EST_HUMAN	601509586F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3911181 5'
8542	21234	34377	0.84	1.5E+00	K02138.1	NT	Mouse germline IgM chain gene, mu-delta region
8914	21605		0.48	1.5E+00	AB038516.1	NT	Homo sapiens hGPIb alpha gene for platelet glycoprotein Ib alpha, complete cds
9032	21722	34876	0.46	1.5E+00	BF217818.1	EST_HUMAN	601882662F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4095135 5'
9383	22045	35217	0.64	1.5E+00	R81928.1	EST_HUMAN	yj03h01.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:147697 5'
9535	22188	35374	1.39	1.5E+00	AW375697.1	EST_HUMAN	QV3-CT0192-261099-008-d09 CT0192 Homo sapiens cDNA
9760	22411	35618	6.39	1.5E+00	BF376754.1	EST_HUMAN	RC0-TN0078-150900-034-g05 TN0078 Homo sapiens cDNA
9952	22600		1.77	1.5E+00	BF337944.1	EST_HUMAN	602035771F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183865 5'
10096	22744	35958	1.66	1.5E+00	AA017689.1	EST_HUMAN	ze38g06.r1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:361306 5'

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10096	22744	35959	1.66	1.5E+00	AA017689.1	EST_HUMAN	ze38g06.r1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:361306 5'
11375	23982	37282	4.46	1.5E+00	AL134197.1	EST_HUMAN	DKFZp547P243_s1 547 (synonym: hfbr1) Homo sapiens cDNA clone DKFZp547P243 3'
11530	24130		6.55	1.5E+00	X07380.1	NT	Maize mitochondrial tRNA-Ser gene and tRNA-Phe pseudogene
11629	24226	37549	2.1	1.5E+00	AI400798.1	EST_HUMAN	tg94d09.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2116433 3'
11629	24226	37550	2.1	1.5E+00	AI400798.1	EST_HUMAN	tg94d09.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2116433 3'
12222	25325	30713	1.44	1.5E+00	D63480.1	NT	Human mRNA for KIAA0146 gene, partial cds
12445	24815		3.38	1.5E+00	AL445065.1	NT	Thermoplasma acidophilum complete genome; segment 3/5
28	12856	25472	2.76	1.4E+00	7661685	NT	Homo sapiens DKFZP586M0122 protein (DKFZP586M0122), mRNA
28	12856	25473	2.76	1.4E+00	7661685	NT	Homo sapiens DKFZP586M0122 protein (DKFZP586M0122), mRNA
2333	15057		6.92	1.4E+00	U67922.1	NT	Ovis aries prion protein gene, complete cds
2675	15384	28125	2.21	1.4E+00	X74463.1	NT	Human papillomavirus type 7 genomic DNA
2776	15481	28221	2.61	1.4E+00	AF064564.2	NT	Fugu rubripes neurofibromatosis type 1 (NF1), A-kinase anchor protein (AKAP84), BAW protein (BAW), and WSB1 protein (WSB1) genes, complete cds
2776	15481	28222	2.61	1.4E+00	AF064564.2	NT	Fugu rubripes neurofibromatosis type 1 (NF1), A-kinase anchor protein (AKAP84), BAW protein (BAW), and WSB1 protein (WSB1) genes, complete cds
4545	17280		1.81	1.4E+00	BF681547.1	EST_HUMAN	602156687F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4297556 5'
5288	18093	30754	1.61	1.4E+00	AW054976.1	EST_HUMAN	wt45g07.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2510460 3'
5441	18240		5.57	1.4E+00	AB032983.1	NT	Homo sapiens mRNA for KIAA1157 protein, partial cds
6186	18963	31936	2.72	1.4E+00	Q13472	SWISSPROT	DNA TOPOISOMERASE III ALPHA
6202	25420		4.02	1.4E+00	AB020712.1	NT	Homo sapiens mRNA for KIAA0905 protein, complete cds
6318	19089	32074	2.67	1.4E+00	Q92777	SWISSPROT	SYNAPSIN II
6318	19089	32075	2.67	1.4E+00	Q92777	SWISSPROT	SYNAPSIN II
7186	19872	32946	2.07	1.4E+00	AJ133269.1	NT	Homo sapiens caveolin-1/-2 locus, Contig1, D7S522, genes CAV2 (exons 1, 2a, and 2b), CAV1 (exons 1 and 2)
7201	19887	32962	1.17	1.4E+00	AW467760.1	EST_HUMAN	he23f05.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2919873 3' similar to contains Alu repetitive element;
7258	19942	33018	0.75	1.4E+00	P55268	SWISSPROT	LAMININ BETA-2 CHAIN PRECURSOR (S-LAMININ)
7258	19942	33019	0.75	1.4E+00	P55268	SWISSPROT	LAMININ BETA-2 CHAIN PRECURSOR (S-LAMININ)
8233	20927		0.68	1.4E+00	P07683	SWISSPROT	GLUCOAMYLASE PRECURSOR (GLUCAN 1,4-ALPHA-GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE)
8693	21385		4.47	1.4E+00	AJ271735.1	NT	Homo sapiens Xq pseudoautosomal region; segment 1/2
8991	21681	34829	1.73	1.4E+00	R20459.1	EST_HUMAN	yg33f12.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:34345 5'
9097	21785	34951	4.65	1.4E+00	BE064667.1	EST_HUMAN	RC1-BT0313-301299-012-f05 BT0313 Homo sapiens cDNA
9131	21819	34985	0.51	1.4E+00	AF134844.1	NT	Sceloporus undulatus ornithine transcarbamylase (OTC) mRNA, complete cds

Page 20 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10108	22756	35968	0.79	1.4E+00	BF575545.1	EST_HUMAN	602133135F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4288137 5'
10151	22799	36015	0.61	1.4E+00	BE145374.1	EST_HUMAN	IL5-HT0198-291099-008-C04 HT0198 Homo sapiens cDNA
10151	22799	36016	0.61	1.4E+00	BE145374.1	EST_HUMAN	IL5-HT0198-291099-008-C04 HT0198 Homo sapiens cDNA
10424	23070	36291	1.06	1.4E+00	D63441.1	NT	Pandorina colemaniae chloroplast rbcL gene for ribulose bisphosphate carboxylase, partial cds
10424	23070	36292	1.06	1.4E+00	D63441.1	NT	Pandorina colemaniae chloroplast rbcL gene for ribulose bisphosphate carboxylase, partial cds
11003	23675	36931	1.34	1.4E+00	AA195528.1	EST_HUMAN	zr36e09.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:665512 5' similar to contains element MER22 repetitive element ;
11188	23853	37139	6.16	1.4E+00	AB006682.1	NT	Homo sapiens APECED mRNA for AIRE-1, complete cds
11381	23988	37288	4.42	1.4E+00	BE962107.2	EST_HUMAN	601655184R1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3845805 3'
11381	23988	37289	4.42	1.4E+00	BE962107.2	EST_HUMAN	601655184R1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3845805 3'
11404	24053	37357	3.46	1.4E+00	U30790.1	NT	Pneumocystis carinii f. sp. ratti guanine nucleotide binding protein alpha subunit (pcg1) gene, complete cds
11404	24053	37358	3.46	1.4E+00	U30790.1	NT	Pneumocystis carinii f. sp. ratti guanine nucleotide binding protein alpha subunit (pcg1) gene, complete cds
12079	25256		1.48	1.4E+00	AL161500.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 12
557	13339		1.81	1.3E+00	Z73640.1	NT	M.mucedo gene encoding 4-Dihydromethyl-trisporate dehydrogenase
882	13651	26320	3.42	1.3E+00	AJ271192.1	NT	Cantharellus sp. partial 25S rRNA gene, isolate Tibet
1107	13864		20.26	1.3E+00	Y19213.1	NT	Homo sapiens putative psihHbA pseudogene for hair keratin, exons 2 to 7
1274	14024	26692	13.71	1.3E+00	4507998	NT	Homo sapiens zinc finger protein 157 (HZF22) (ZNF157) mRNA
1274	14024	26693	13.71	1.3E+00	4507998	NT	Homo sapiens zinc finger protein 157 (HZF22) (ZNF157) mRNA
1334	14083		1.26	1.3E+00	U61730.2	NT	Coix lacryma-jobi dihydrodipicolinate synthase (dapA) gene, complete cds
1605	14351		2.27	1.3E+00	AE002338.2	NT	Chlamydia muridarum, section 66 of 85 of the complete genome
2239	14967		1	1.3E+00	AB030447.1	NT	Cyprinus carpio MRPb and MASPb genes for mannose-binding lectin-associated serine protease (MASP) and MASP-related protein, complete cds
2405	15126	27862	1.27	1.3E+00	P25391	SWISSPROT	LAMININ ALPHA-1 CHAIN PRECURSOR (LAMININ A CHAIN)
2553	15268		1.75	1.3E+00	BE966735.2	EST_HUMAN	601661233R1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3915945 3'
2940	15705	28354	0.73	1.3E+00	6755621	NT	Mus musculus alpha-spectrin 1, erythroid (Spna1), mRNA
3584	16339	28984	0.89	1.3E+00	AF016494.1	NT	Fugu rubripes gamma-aminobutyric acid receptor beta subunit gene, partial cds; 55kd erythrocyte membrane protein (P55), synaptic vesicle-associated integral membrane protein (VAMP-1), procollagen C-proteinase enhancer protein (PCOLCE) genes, complete c>
5427	18226	30938	1.09	1.3E+00	P19732	SWISSPROT	PHENOL HYDROXYLASE P3 PROTEIN (PHENOL 2-MONOOXYGENASE P3 COMPONENT)
5622	18418	31330	0.6	1.3E+00	M27138.1	NT	Human estradiol 17 beta-dehydrogenase gene, complete cds
5863	18650	31590	0.81	1.3E+00	BF663825.1	EST_HUMAN	602145264F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4309095 5'
5928	18712	31667	7.57	1.3E+00	AW362834.1	EST_HUMAN	PM0-CT0289-291199-004-f08 CT0289 Homo sapiens cDNA

Page 21 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5928	18712	31668	7.57	1.3E+00	AW362834.1	EST_HUMAN	PM0-CT0289-291199-004-f08 CT0289 Homo sapiens cDNA
6323	19093	32081	1.34	1.3E+00	M33496.1	NT	D.melanogaster no-on-transient A gene product, complete cds
6652	19414		0.76	1.3E+00	Q00156	SWISSPROT	HYPOTHETICAL GENE 64 PROTEIN
6739	19573	32606	0.62	1.3E+00	M13918.2	NT	Homo sapiens fibronectin receptor alpha-subunit precursor (ITGA5) mRNA, partial cds
6854	19554	32584	1.17	1.3E+00	BE538819.1	EST_HUMAN	601061420F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3447965 5'
7000	19692	32743	0.81	1.3E+00	BE243571.1	EST_HUMAN	TCBAP1D0959 Pediatric pre-B cell acute lymphoblastic leukemia Baylor-HGSC project=TCBA Homo sapiens cDNA clone TCBAP0959
7358	20039	33117	1.01	1.3E+00	P24540	SWISSPROT	ACYLPHOSPHATASE, ORGAN-COMMON TYPE ISOZYMES A AND B (ACYLPHOSPHATE PHOSPHOHYDROLASE)
8197	20891	34029	1.28	1.3E+00	AJ009912.1	NT	Sus scrofa pip gene
8346	21039	34176	2.78	1.3E+00	BE963379.2	EST_HUMAN	601657145R1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3866195 3'
8459	21151	34294	0.86	1.3E+00	BE974280.1	EST_HUMAN	601680250R2 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:3950532 3'
8611	21303		1.78	1.3E+00	9910247	NT	Homo sapiens GL004 protein (GL004), mRNA
8689	21381	34525	0.79	1.3E+00	AI927629.1	EST_HUMAN	wo85a07.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2462100 3'
9415	22093		5.24	1.3E+00	AF042084.1	NT	Homo sapiens heparan glucosaminyl N-deacetylase/N-sulfotransferase-2 gene, complete cds
9424	22102	35273	2.56	1.3E+00	X72019.1	NT	S.alba phr-1 mRNA for photolyase
9424	22102	35274	2.56	1.3E+00	X72019.1	NT	S.alba phr-1 mRNA for photolyase
9524	22177	35361	0.86	1.3E+00	AF059250.1	NT	Homo sapiens lipoxygenase (ALOX12B) mRNA, complete cds
9569	22222	35407	1.56	1.3E+00	O00754	SWISSPROT	LYSOSOMAL ALPHA-MANNOSIDASE PRECURSOR (MANNOSIDASE, ALPHA B) (LYSOSOMAL ACID ALPHA-MANNOSIDASE) (LAMAN)
9651	22303	35498	1.14	1.3E+00	AI927629.1	EST_HUMAN	wo85a07.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2462100 3'
9726	22377	35578	0.79	1.3E+00	AJ223962.1	NT	Lactococcus lactis cremoris NCDO-inv1 chromosomal inversion junction DNA
9726	22377	35579	0.79	1.3E+00	AJ223962.1	NT	Lactococcus lactis cremoris NCDO-inv1 chromosomal inversion junction DNA
9766	22417	35624	4.53	1.3E+00	BE963379.2	EST_HUMAN	601657145R1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3866195 3'
9826	22477		0.48	1.3E+00	AI559944.1	EST_HUMAN	tq77a12.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2214814 3' similar to gb:X14723 CLUSTERIN PRECURSOR (HUMAN);
10050	22698	35913	0.46	1.3E+00	AF061251.1	NT	Eschericia coli serotype O157:H7 O antigen gene cluster
10050	22698	35914	0.46	1.3E+00	AF061251.1	NT	Eschericia coli serotype O157:H7 O antigen gene cluster
10113	22761	35974	1.62	1.3E+00	AE004392.1	NT	Vibrio cholerae chromosome II, section 49 of 93 of the complete chromosome
10130	22778	35991	1.36	1.3E+00	M29953.1	NT	Campylobacter jejuni kanamycin phosphotransferase (aphA-7) gene, complete cds
10483	23129		0.82	1.3E+00	AL163302.2	NT	Homo sapiens chromosome 21 segment HS21C102
10511	23157	36383	0.45	1.3E+00	AI990846.1	EST_HUMAN	ws32e10.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2498922 3' similar to SW:TRXB_HUMAN Q16881 THIOREDOXIN REDUCTASE ;
10592	23286		4.6	1.3E+00	Q14117	SWISSPROT	DIHYDROPYRIMIDINASE (DHPASE) (HYDANTOINASE) (DHP)

Page 22 of 536

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10822	23505	36744	1.93	1.3E+00	P25299	SWISSPROT	MRNA 3'-END PROCESSING PROTEIN RNA15
10846	23528	36772	2.01	1.3E+00	Z18892.2	NT	Mus musculus desmin gene
11307	23966		1.8	1.3E+00	AW274791.1	EST_HUMAN	xp09e03.x1 NCI_CGAP_HN9 Homo sapiens cDNA clone IMAGE:2739868 3'
11527	24127	37433	3.21	1.3E+00	D42042.1	NT	Human mRNA for KIAA0085 gene, partial cds
11624	24221	37544	3.16	1.3E+00	Z98682.1	NT	Bacillus subtilis genomic DNA 23.9kB fragment
12210	24675		2.64	1.3E+00	AF187873.1	NT	Cavia porcellus inwardly-rectifying potassium channel Kir2.2 (KCNJ12) gene, complete cds
12388	24780	31035	6.3	1.3E+00	BF348043.1	EST_HUMAN	602023185F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4158452 5'
12397	25153		2.73	1.3E+00	P33464	SWISSPROT	E1 GLYCOPROTEIN PRECURSOR (MATRIX GLYCOPROTEIN) (MEMBRANE GLYCOPROTEIN)
12489	24848		2.15	1.3E+00	AF187035.1	NT	Sturnira lilium cytochrome b gene, complete cds; mitochondrial gene for mitochondrial product
635	13414	26050	11.05	1.2E+00	AA676246.1	EST_HUMAN	zi22d08.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:431535 3'
804	13576	26239	0.87	1.2E+00	P05228	SWISSPROT	HISTIDINE-RICH PROTEIN PRECURSOR (CLONE PFHRP-III)
804	13576	26240	0.87	1.2E+00	P05228	SWISSPROT	HISTIDINE-RICH PROTEIN PRECURSOR (CLONE PFHRP-III)
804	13576	26241	0.87	1.2E+00	P05228	SWISSPROT	HISTIDINE-RICH PROTEIN PRECURSOR (CLONE PFHRP-III)
858	13627		1.35	1.2E+00	8924234	NT	Homo sapiens hypothetical protein PRO3077 (PRO3077), mRNA
1138	13893	26554	5.64	1.2E+00	AF080245.2	NT	Elaeis oleifera sesquiterpene synthase mRNA, complete cds
1183	13935	26600	1.26	1.2E+00	AJ252242.1	NT	pea seed-borne mosaic virus complete genome
1183	13935	26601	1.26	1.2E+00	AJ252242.1	NT	pea seed-borne mosaic virus complete genome
2003	14739	27463	1.22	1.2E+00	AF140631.1	NT	Homo sapiens G-protein coupled receptor 14 (GPR14) gene, complete cds
3108	15873	28512	1.24	1.2E+00	AB020681.1	NT	Homo sapiens mRNA for KIAA0874 protein, partial cds
3163	15926	28573	5.98	1.2E+00	AL161563.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 63
3163	15926	28574	5.98	1.2E+00	AL161563.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 63
3280	16041		2.59	1.2E+00	P54910	SWISSPROT	CONJUGAL TRANSFER PROTEIN TRBE PRECURSOR
3699	16452	29091	6.69	1.2E+00	U75902.1	NT	Mus musculus subtilisin-like serine protease LPC (PC7) gene, exons 1 to 9, partial cds
3967	16716	29354	1.78	1.2E+00	BF373570.1	EST_HUMAN	MR0-FT0175-050900-203-g06_1 FT0175 Homo sapiens cDNA
4266	16110	28766	1.11	1.2E+00	AF188740.1	NT	Homo sapiens LHX3 gene, intron 2
4438	17174		1.57	1.2E+00	M87060.1	NT	Rattus rattus cardiac AE3 gene, exons 1-23
4487	17222	29850	0.96	1.2E+00	AL161509.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 21
4523	17258	29892	1.89	1.2E+00	AF156495.1	NT	Homo sapiens post-synaptic density 95 (DLG4) gene, complete cds
4548	17283		6.43	1.2E+00	Y09200.1	NT	T.pinnatum chloroplast rbcL gene, partial
5351	18154	30836	1.1	1.2E+00	U20760.1	NT	Human extracellular calcium-sensing receptor mRNA, complete cds
5467	18266	31158	1.91	1.2E+00	AW813276.1	EST_HUMAN	MR3-ST0191-140200-013-c05 ST0191 Homo sapiens cDNA
5784	18575	31504	0.83	1.2E+00	AF016052.1	NT	Homo sapiens zinc finger protein ZNF191 (ZNF191) gene, complete cds
6060	18840	31801	2.51	1.2E+00	X74885.1	NT	D.hydei ay1 repeat cluster DNA, fragment D
6119	18897	31865	4.42	1.2E+00	BE003113.1	EST_HUMAN	QV4-BN0090-270400-190-a03 BN0090 Homo sapiens cDNA

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6198	18974	31951	1.54	1.2E+00	X89084.1	NT	C.glutamicum pta gene and ackA gene
6198	18974	31952	1.54	1.2E+00	X89084.1	NT	C.glutamicum pta gene and ackA gene
6241	19015	31989	39.54	1.2E+00	AA759254.1	EST_HUMAN	ah84g12.s1 Soares_testis_NHT Homo sapiens cDNA clone 1322374 3'
6342	19112	32101	0.55	1.2E+00	N33295.1	EST_HUMAN	yy39b12.s1 Soares melanocyte 2NbHM Homo sapiens cDNA clone IMAGE:273599 3' similar to gb|M87935|HUMAALU472 Human carcinoma cell-derived Alu RNA transcript, (rRNA); gb:J04970 CARBOXYPEPTIDASE M PRECURSOR (HUMAN);
6408	19177	32175	0.68	1.2E+00	P17671	SWISSPROT	ECDYSONE-INDUCIBLE PROTEIN E75-A
6412	19180	32179	2.06	1.2E+00	AW813276.1	EST_HUMAN	MR3-ST0191-140200-013-c05 ST0191 Homo sapiens cDNA
6815	19476	32498	1.17	1.2E+00	AB029010.1	NT	Homo sapiens mRNA for KIAA1087 protein, partial cds
6829	19490	32512	3.11	1.2E+00	AJ002141.1	NT	Mus musculus DSPP gene
7153	19840		0.94	1.2E+00	AJ271735.1	NT	Homo sapiens Xq pseudoautosomal region; segment 1/2
7282	25109	33044	4.86	1.2E+00	AV734585.1	EST_HUMAN	AV734585 cdA Homo sapiens cDNA clone cdAAFH03 5'
7550	20220	33323	2.49	1.2E+00	X74207.1	NT	L.lactis pyrD and pyrF genes
7603	20269	33376	0.56	1.2E+00	J05218.1	NT	Chicken muscarinic acetylcholine receptor (cm4 mAChR) gene, complete cds
7715	20379	33492	0.56	1.2E+00	BE787646.1	EST_HUMAN	601481761F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3884270 5'
8467	21159	34302	3.32	1.2E+00	AB033030.1	NT	Homo sapiens mRNA for KIAA1204 protein, partial cds
8561	21253	34391	0.68	1.2E+00	P38427	SWISSPROT	ALPHA,ALPHA-TREHALOSE-PHOSPHATE SYNTHASE [UDP-FORMING] 123 KD SUBUNIT (TREHALOSE-6-PHOSPHATE SYNTHASE) (UDP-GLUCOSE-GLUCOSEPHOSPHATE GLUCOSYLTRANSFERASE)
8775	21467		0.51	1.2E+00	7706271	NT	Homo sapiens CGI-30 protein (LOC51611), mRNA
8923	21614	34758	1.87	1.2E+00	AW377210.1	EST_HUMAN	MR2-CT0222-201099-001-e07 CT0222 Homo sapiens cDNA
9138	21826	34991	0.5	1.2E+00	H48599.1	EST_HUMAN	yq80a06.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:202066 5'
9298	21985	35138	3.75	1.2E+00	Z32850.1	NT	R.communis gene for pyrophosphate-dependent phosphofructokinase beta subunit
9505	22158	35339	1.81	1.2E+00	D11745.1	EST_HUMAN	HUMHM01A01 Liver HepG2 cell line. Homo sapiens cDNA clone hm01a01
9831	22482	35684	2.86	1.2E+00	X56832.1	NT	H.sapiens ENO3 gene for muscle specific enolase
10224	22872		0.73	1.2E+00	AB009666.1	NT	Homo sapiens klotho gene, exon 1
11318	24009	37314	3.78	1.2E+00	AW817817.1	EST_HUMAN	PM0-ST0264-161199-001-d01 ST0264 Homo sapiens cDNA
11357	24045		10.62	1.2E+00	BE160761.1	EST_HUMAN	PM1-HT0422-160200-007-g10 HT0422 Homo sapiens cDNA
11435	23202	36434	4.36	1.2E+00	U50147.1	NT	Rattus norvegicus synapse-associated protein 102 mRNA, complete cds
12179	25227	30817	17.06	1.2E+00	AL163203.2	NT	Homo sapiens chromosome 21 segment HS21C003
12199	24667		2.8	1.2E+00	AP001515.1	NT	Bacillus halodurans genomic DNA, section 9/14
451	13237	25876	1.53	1.1E+00	D86980.1	NT	Human mRNA for KIAA0227 gene, partial cds
1757	14499	27200	1.33	1.1E+00	AW995393.1	EST_HUMAN	QV0-BN0042-170300-163-g12 BN0042 Homo sapiens cDNA
1892	14629	27339	0.88	1.1E+00	AW575889.1	EST_HUMAN	UI-HF-BR0p-ajk-f-02-0-UI.s1 NIH_MGC_52 Homo sapiens cDNA clone IMAGE:3074834 3'

Page 24 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
3324	16084	28734	6.48	1.1E+00	AL163213.2	NT	Homo sapiens chromosome 21 segment HS21C013
3324	16084	28735	6.48	1.1E+00	AL163213.2	NT	Homo sapiens chromosome 21 segment HS21C013
3480	16236	28892	1.11	1.1E+00	8922641	NT	Homo sapiens hypothetical protein FLJ10749 (FLJ10749), mRNA
3567	16322	28970	1.01	1.1E+00	AI808360.1	EST_HUMAN	wf54h11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2359461 3' similar to SW:P531_HUMAN Q12888 P53-BINDING PROTEIN 53BP1 ;
3707	16460	29098	1.05	1.1E+00	AE003886.1	NT	Xylella fastidiosa, section 32 of 229 of the complete genome
3707	16460	29099	1.05	1.1E+00	AE003886.1	NT	Xylella fastidiosa, section 32 of 229 of the complete genome
3798	16550		1.02	1.1E+00	X85374.1	NT	H.parahaemolyticus hphIM(A), hphIM(C), hphIR and menB genes
4190	16931		5.69	1.1E+00	5835331	NT	R.unicornis complete mitochondrial genome
4634	17369		0.81	1.1E+00	U34992.1	NT	Carcharhinus plumbeus Ig lambda light chain gene, complete cds
4934	17682	30272	3.45	1.1E+00	U18466.1	NT	African swine fever virus, complete genome
4935	17683	30273	1.05	1.1E+00	AJ271740.1	NT	Drosophila melanogaster D-Titin gene, exons 1-37
5129	17847	30464	1.07	1.1E+00	6680080	NT	Homo sapiens putative GR6 protein (GR6), mRNA
5224	18031	30657	1.39	1.1E+00	6978530	NT	Rattus norvegicus Aquaporin 4 (Aqp4), mRNA
5526	18324	31225	15.75	1.1E+00	BE960184.1	EST_HUMAN	601652776R1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3825835 3'
5545	18342	31250	1.2	1.1E+00	AI138582.1	EST_HUMAN	qd85c03.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1736260 3'
6001	18782	31743	1.1	1.1E+00	11419739	NT	Homo sapiens solute carrier family 6 (neurotransmitter transporter), member 14 (SLC6A14), mRNA
6181	18958	31932	0.62	1.1E+00	AF197861.1	NT	Macgregoria pulchra cytochrome b gene, complete cds; mitochondrial gene for mitochondrial product
6313	19084	32069	0.82	1.1E+00	R06037.1	EST_HUMAN	ye89e03.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:124924 5'
6616	19379	32394	0.72	1.1E+00	AJ404004.1	NT	Mus musculus mRNA for ER protein 58 (EP58 gene)
7155	19842		0.58	1.1E+00	AF101091.1	NT	Homo sapiens collagen type XI alpha-1 (COL11A1) gene, exons 25 through 28
7196	19882	32956	0.72	1.1E+00	X55981.1	NT	Maize mRNA for enolase (2-phospho-D-glycerate hydrolase)
7389	20068	33146	2.18	1.1E+00	Z72338.1	NT	Herpes simplex virus type 1 (strain KOS) UL41 gene
7389	20068	33147	2.18	1.1E+00	Z72338.1	NT	Herpes simplex virus type 1 (strain KOS) UL41 gene
7411	20088	33172	8.84	1.1E+00	AL161588.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 84
7480	25115	33247	0.8	1.1E+00	11967960	NT	Mus musculus silent mating type information regulation 2, (S.cerevisiae, homolog)-like (Sir2l), mRNA
8032	20727	33860	3.01	1.1E+00	BF693996.1	EST_HUMAN	602082582F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4246828 5'
8120	20814	33950	0.64	1.1E+00	AI478339.1	EST_HUMAN	tm39h11.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2160549 3'
8636	21328	34471	0.71	1.1E+00	AB003088.1	NT	Acetabularia caliculus mitochondrial COXI-like gene
8714	21406	34549	0.75	1.1E+00	S80750.1	NT	VH=anti-cytomegalovirus glycoprotein B antibody 4D4 heavy chain variable region [human, mRNA Partial, 375 nt]

Page 25 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8824	21516	34661	0.45	1.1E+00	AI079946.1	EST_HUMAN	oz34f05.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1677249 3'
9337	20408		0.69	1.1E+00	BE384876.1	EST_HUMAN	601276278F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3617418 5'
9528	22181	35365	0.53	1.1E+00	AJ245772.1	NT	Mus musculus mRNA for stretch responsive muscle (X-chromosome) protein (Smx gene)
9580	22233		1.2	1.1E+00	Y12227.1	NT	Arabidopsis thaliana DNA, 24 kb surrounding PFL locus
9672	22324	35520	1.14	1.1E+00	L76301.1	NT	Yersinia pseudotuberculosis psaE, psaF, adhesin (psaA), chaperone (psaB), and usher (psaC) genes, complete cds
9732	22383	35585	1.37	1.1E+00	AB023151.1	NT	Homo sapiens mRNA for KIAA0934 protein, partial cds
9837	22488	35690	4.59	1.1E+00	AL161515.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 27
9898	22548	35742	18.34	1.1E+00	6754021	NT	Mus musculus guanine nucleotide binding protein (G protein), gamma 3 subunit (Gng3), mRNA
10398	23044	36260	1.1	1.1E+00	P73769	SWISSPROT	DNA MISMATCH REPAIR PROTEIN MUTS
10504	23150	36375	0.73	1.1E+00	AI878921.1	EST_HUMAN	au51c11.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone IMAGE:2518292 5' similar to gb:D10522 Human mRNA for 80K-L protein, complete cds. (HUMAN);
10547	23243	36478	2.25	1.1E+00	11067364	NT	Homo sapiens KIAA0626 gene product (KIAA0626), mRNA
10606	23300		3.1	1.1E+00	AF068942.1	NT	Klebsormidium flultans cytochrome c oxidase subunit 2 (cox2) gene, mitochondrial gene encoding mitochondrial protein, partial cds
11023	23695	36958	1.28	1.1E+00	11439596	NT	Homo sapiens potassium inwardly-rectifying channel, subfamily J, member 11 (KCNJ11), mRNA
11026	23698	36961	1.58	1.1E+00	L16877.1	NT	Homo sapiens cytochrome P4502C9 (CYP2C9) gene, 5' flank and exon 1
11042	17901		5.23	1.1E+00	8922973	NT	Homo sapiens hypothetical protein FLJ11280 (FLJ11280), mRNA
11048	23718	36988	3.68	1.1E+00	AF012862.1	NT	Petroselinum crispum cytosolic glucose-6-phosphate dehydrogenase 1 (cG6PDH1) mRNA, complete cds
11048	23718	36989	3.68	1.1E+00	AF012862.1	NT	Petroselinum crispum cytosolic glucose-6-phosphate dehydrogenase 1 (cG6PDH1) mRNA, complete cds
11328	24019	37323	4.58	1.1E+00	AI809699.1	EST_HUMAN	wf76e11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2361548 3'
11561	24160	37470	1.63	1.1E+00	D89501.1	NT	Human PBI gene, complete cds
11561	24160	37471	1.63	1.1E+00	D89501.1	NT	Human PBI gene, complete cds
12153	24639		3.66	1.1E+00	P07866	SWISSPROT	LOW TEMPERATURE ESSENTIAL PROTEIN
12250	24697	31078	1.93	1.1E+00	AF216698.1	NT	Taenia solium immunogenic protein Ts76 mRNA, partial cds
12378	25225		2.09	1.1E+00	AF234169.1	NT	Dictyostelium discoideum isopentenyl pyrophosphate isomerase (Dipi) mRNA, complete cds
12388	25200		1.44	1.1E+00	8393196	NT	Rattus norvegicus C-reactive protein, member of the pentraxin family (Crp), mRNA
97	12923		2.46	1.0E+00	U23808.1	NT	Xenopus laevis rhodopsin gene, complete cds
111	12932	25569	0.73	1.0E+00	D88425.1	NT	Cavia cobaya mRNA for serine/threoine kinase, complete cds
409	13194		2.25	1.0E+00	AB021684.1	NT	Marchantia polymorpha genes for 26S rRNA, 5S rRNA, 18S rRNA, 5.8S rRNA and 26S rRNA
562	13344	25971	1.2	1.0E+00	AJ251660.1	NT	Girardia tigrina mRNA for homeodomain transcription factor (so gene)
662	13438	26079	4.38	1.0E+00	AL163218.2	NT	Homo sapiens chromosome 21 segment HS21C018

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
663	13439		0.95	1.0E+00	AF125984.1	NT	Aedes aegypti mucin-like protein MUC1 mRNA, complete cds
1365	15567		3.03	1.0E+00	X80416.1	NT	V.carteri Algal-CAM mRNA
1751	14493	27193	0.93	1.0E+00	AB006531.1	NT	Plautia stali intestine virus RNA for nonstructural polyprotein, capsid protein precursor, complete cds
2489	15206	27947	1.18	1.0E+00	P48355	SWISSPROT	DNA GYRASE SUBUNIT B
2489	15206	27948	1.18	1.0E+00	P48355	SWISSPROT	DNA GYRASE SUBUNIT B
2878	15645	28287	3.82	1.0E+00	P24008	SWISSPROT	3-OXO-5-ALPHA-STEROID 4-DEHYDROGENASE 1 (STEROID 5-ALPHA-REDUCTASE 1) (SR TYPE 1)
2878	15645	28288	3.82	1.0E+00	P24008	SWISSPROT	3-OXO-5-ALPHA-STEROID 4-DEHYDROGENASE 1 (STEROID 5-ALPHA-REDUCTASE 1) (SR TYPE 1)
2967	15733		1.17	1.0E+00	O14226	SWISSPROT	HYPOTHETICAL 67.9 KD PROTEIN C6F12.08C IN CHROMOSOME I
3194	15957	28609	1.24	1.0E+00	AA628453.1	EST_HUMAN	af28g08.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:1032830 3' similar to WP:C42D8.3 CE04204 ;contains element MER22 MER22 repetitive element ;
3585	12923		1.24	1.0E+00	U23808.1	NT	Xenopus laevis rhodopsin gene, complete cds
3669	16422	29063	1.04	1.0E+00	AJ223816.1	NT	Agaricus bisporus mRNA for tyrosinase
4050	16795	29424	0.76	1.0E+00	AF223391.1	NT	Homo sapiens calcium channel alpha1E subunit (CACNA1E) gene, exons 7-49, and partial cds, alternatively spliced
4242	16983		0.79	1.0E+00	8922245	NT	Homo sapiens hypothetical protein FLJ10139 (FLJ10139), mRNA
4954	17680		0.93	1.0E+00	D10852.1	NT	Rattus norvegicus mRNA for N-acetylglucosaminyltransferase III, complete cds
4975	17698	30306	0.74	1.0E+00	AF092505.1	NT	Mus musculus dipeptidyl aminopeptidase-like protein 6 (Dpp6) gene, partial cds; and proximal Rump white inversion breakpoint
5200	18008	30629	3.53	1.0E+00	Z97022.1	NT	Hordeum vulgare gene encoding cysteine proteinase
5759	18551	31472	4.97	1.0E+00	AF248054.1	NT	Bos taurus micromolar calcium activated neutral protease 1 (CAPN1) gene, exons 11-20, and partial cds
5759	18551	31473	4.97	1.0E+00	AF248054.1	NT	Bos taurus micromolar calcium activated neutral protease 1 (CAPN1) gene, exons 11-20, and partial cds
5867	18654	31595	1.53	1.0E+00	Z97341.2	NT	Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment No. 6
6024	18804	31765	4.7	1.0E+00	P04501	SWISSPROT	FIBER PROTEIN
6030	18810	31770	1.49	1.0E+00	AW452782.1	EST_HUMAN	UI-H-BI3-alx-d-09-0-UI.s1 NCI_CGAP_Sub5 Homo sapiens cDNA clone IMAGE:3068969 3'
6397	19166	32166	1.95	1.0E+00	U75902.1	NT	Mus musculus subtilisin-like serine protease LPC (PC7) gene, exons 1 to 9, partial cds
6447	19215	32213	0.91	1.0E+00	AF104669.1	NT	Homo sapiens cell cycle protein (PA2G4) gene, exons 2 though 5
6534	19300		1.06	1.0E+00	P46506	SWISSPROT	SRB-11 PROTEIN
6679	19596	32634	1.33	1.0E+00	Y11204.1	NT	V.carteri gene encoding volvoxopsin

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7038	19730	32788	1.09	1.0E+00	S52770.1	NT	Insulin-like growth factor-binding protein 4 [cattle, pulmonary artery endothelial cells, mRNA, 2028 nt]
7378	20058		9.29	1.0E+00	P20273	SWISSPROT	B-CELL RECEPTOR CD22 PRECURSOR (LEU-14) (B-LYMPHOCYTE CELL ADHESION MOLECULE) (BL-CAM)
7611	20277	33385	1.56	1.0E+00	AF192531.1	NT	Homo sapiens endothelin-converting enzyme 2 (ECE2) mRNA, complete cds
7626	20292	33401	5.26	1.0E+00	AA775191.1	EST_HUMAN	ac79b08.s1 Stratagene lung (#937210) Homo sapiens cDNA clone IMAGE:868791 3'
7861	20556	33681	1.36	1.0E+00	BE868267.1	EST_HUMAN	601443950F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3848005 5'
7861	20556	33682	1.36	1.0E+00	BE868267.1	EST_HUMAN	601443950F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3848005 5'
8041	17680		1.19	1.0E+00	D10852.1	NT	Rattus norvegicus mRNA for N-acetylglucosaminyltransferase III, complete cds
8248	20942	34079	2.02	1.0E+00	Q02207	SWISSPROT	PEROXISOMAL HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE) (MULTIFUNCTIONAL BETA-OXIDATION PROTEIN) (MFP) [INCLUDES: 2-ENOYL-COA HYDRATASE ; D-3-HYDROXYACYL COA DEHYDROGENASE]
8248	20942	34080	2.02	1.0E+00	Q02207	SWISSPROT	PEROXISOMAL HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE) (MULTIFUNCTIONAL BETA-OXIDATION PROTEIN) (MFP) [INCLUDES: 2-ENOYL-COA HYDRATASE ; D-3-HYDROXYACYL COA DEHYDROGENASE]
8376	21069		0.85	1.0E+00	P51784	SWISSPROT	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 11 (UBIQUITIN THIOLESTERASE 11) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 11) (DEUBIQUITINATING ENZYME 11)
8408	21101	34237	0.5	1.0E+00	Q9Y5T5	SWISSPROT	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 16 (UBIQUITIN THIOLESTERASE 16) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 16) (DEUBIQUITINATING ENZYME 16) (UBIQUITIN PROCESSING PROTEASE UBP-M)
8408	21101	34238	0.5	1.0E+00	Q9Y5T5	SWISSPROT	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 16 (UBIQUITIN THIOLESTERASE 16) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE 16) (DEUBIQUITINATING ENZYME 16) (UBIQUITIN PROCESSING PROTEASE UBP-M)
8436	25122		2.34	1.0E+00	BE147331.1	EST_HUMAN	RC1-HT0229-181099-011-e06 HT0229 Homo sapiens cDNA
8476	21168	34312	0.88	1.0E+00	U42720.2	NT	Simian immunodeficiency virus Gag protein (gag) gene, complete cds; Pol protein (pol) gene, partial cds; and Vif protein (vif), Vpr protein (vpr), Tat protein (tat), Rev protein (rev), Vpu protein (vpu), Env protein (env), and Nef protein (nef) genes, >
8625	21317	34459	1.27	1.0E+00	M38427.1	NT	Human immunodeficiency virus type 1 (HIV-1), isolate SF33,
9171	21841	35006	2.43	1.0E+00	BE907592.1	EST_HUMAN	601497581F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3899421 5'
9381	22043	35215	1.69	1.0E+00	6753429	NT	Mus musculus chloride channel calcium activated 1 (Clca1), mRNA
9381	22043	35216	1.69	1.0E+00	6753429	NT	Mus musculus chloride channel calcium activated 1 (Clca1), mRNA
9510	22163	35345	1.83	1.0E+00	AV689554.1	EST_HUMAN	AV689554 GKC Homo sapiens cDNA clone GKCCYA11 5'
9516	22169	35351	1.43	1.0E+00	U44952.1	NT	Xenopus laevis zona pellucida C glycoprotein precursor (xlZPC) mRNA, complete cds
9516	22169	35352	1.43	1.0E+00	U44952.1	NT	Xenopus laevis zona pellucida C glycoprotein precursor (xlZPC) mRNA, complete cds

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9753	22404	35609	0.49	1.0E+00	X15498.1	NT	Human Coronavirus gene for membrane protein
9753	22404	35610	0.49	1.0E+00	X15498.1	NT	Human Coronavirus gene for membrane protein
10012	22660	35875	0.71	1.0E+00	5174562	NT	Homo sapiens MHC binding factor, beta (MHCBFB) mRNA
10012	22660	35876	0.71	1.0E+00	5174562	NT	Homo sapiens MHC binding factor, beta (MHCBFB) mRNA
10104	22752	35966	0.81	1.0E+00	AI077920.1	EST_HUMAN	oy15d07.s1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:1665901 3'
10225	22873	36085	4.36	1.0E+00	AV758825.1	EST_HUMAN	AV758825 BM Homo sapiens cDNA clone BMFAWC04 5'
10375	23021	36237	16.16	1.0E+00	AA004982.1	EST_HUMAN	zh94a02.r1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:428906 5'
10375	23021	36238	16.16	1.0E+00	AA004982.1	EST_HUMAN	zh94a02.r1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:428906 5'
10407	23053	36270	1.1	1.0E+00	L11910.1	NT	Human retinoblastoma susceptibility gene exons 1-27, complete cds
10893	23573	36823	4.57	1.0E+00	S90825.1	NT	PBR1=proline-rich protein {intron 3} [human, Genomic, 898 nt]
11025	23697	36960	1.49	1.0E+00	AA701494.1	EST_HUMAN	zi63b11.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:435453 3' similar to contains Alu repetitive element;contains element MER38 repetitive element ;
11522	24122		1.59	1.0E+00	L47613.1	NT	Picea glauca EMB13 mRNA
11744	18008	30629	1.55	1.0E+00	Z97022.1	NT	Hordeum vulgare gene encoding cysteine proteinase
11838	24422	37763	12.29	1.0E+00	Q60019	SWISSPROT	NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 8 (NADH DEHYDROGENASE 1, CHAIN 8) (NDH-1, CHAIN 8)
11868	24452	37794	1.38	1.0E+00	9626187	NT	Human adenovirus type 5, complete genome
12049	24568		3.01	1.0E+00	P15306	SWISSPROT	THROMBOMODULIN PRECURSOR (FETOMODULIN) (TM)
12370	24772		2.32	1.0E+00	AW976184.1	EST_HUMAN	EST388293 MAGE resequences, MAGN Homo sapiens cDNA
2643	15353	28097	1.19	9.9E-01	AL163302.2	NT	Homo sapiens chromosome 21 segment HS21C102
3591	16345		0.97	9.9E-01	AF174585.1	NT	Apple mosaic virus RNA 2 putative polymerase gene, complete cds
5547	18344	31253	10.09	9.9E-01	P49657	SWISSPROT	SERINE/THREONINE PROTEIN KINASE MINIBRAIN
5779	18570	31498	0.93	9.9E-01	Q09632	SWISSPROT	PROBABLE OXIDOREDUCTASE ZK1290.5 IN CHROMOSOME II
9160	21830		1.37	9.9E-01	U65667.1	NT	Lycopersicon esculentum putative Mi1 copy 1 nematode-resistance gene
9455	22005		2.18	9.9E-01	Q28642	SWISSPROT	B2 BRADYKININ RECEPTOR (BK-2 RECEPTOR)
10614	23308	36547	2.37	9.9E-01	AJ005029.1	NT	Danio rerio mRNA for Eph-like receptor tyrosine kinase rtk8
11592	24191	37508	2.3	9.9E-01	Y11972.1	NT	B.aphidicola 16S rDNA (host T.suberi)
11592	24191	37509	2.3	9.9E-01	Y11972.1	NT	B.aphidicola 16S rDNA (host T.suberi)
510	13294	25926	1.14	9.8E-01	P22567	SWISSPROT	AMINO-ACID ACETYLTRANSFERASE (N-ACETYLGLUTAMATE SYNTHASE) (AGS) (NAGS)
2295	15020		1.21	9.8E-01	AJ003108.1	NT	Callithrix jacchus UBE1 gene derived retroposon on the Y chromosome
2804	15509		1.01	9.8E-01	AF174644.1	NT	Xenopus laevis rac GTPase mRNA, complete cds
3781	16533	29171	0.92	9.8E-01	O67551	SWISSPROT	PROBABLE ENDONUCLEASE IV (ENDODEOXYRIBONUCLEASE IV)
7099	19788	32852	4.67	9.8E-01	AJ302158.1	NT	Enterobacteriaceae sp. JM983 partial groES gene for GroES-like protein and partial groEL gene for GroEL-like protein, isolate JM983

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7099	19788	32853	4.67	9.8E-01	AJ302158.1	NT	Enterobacteriaceae sp. JM983 partial groES gene for GroES-like protein and partial groEL gene for GroEL-like protein, isolate JM983
7545	20215	33316	1.15	9.8E-01	BF034016.1	EST_HUMAN	601456337F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3860049 5'
7545	20215	33317	1.15	9.8E-01	BF034016.1	EST_HUMAN	601456337F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3860049 5'
8619	21311	34453	0.91	9.8E-01	P38652	SWISSPROT	PHOSPHOGLUCOMUTASE (GLUCOSE PHOSPHOMUTASE) (PGM)
10336	22983		1.13	9.8E-01	AA825565.1	EST_HUMAN	od55d04.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1371847 3'
10916	23596	36842	2.29	9.8E-01	BE258705.1	EST_HUMAN	601110258F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3350750 5'
10916	23596	36843	2.29	9.8E-01	BE258705.1	EST_HUMAN	601110258F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3350750 5'
11764	24355	37688	1.57	9.8E-01	AI680876.1	EST_HUMAN	tx42c10.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2272242 3'
12256	24702		1.56	9.8E-01	U52111.2	NT	Homo sapiens X28 region near ALD locus containing dual specificity phosphatase 9 (DUSP9), ribosomal protein L18a (RPL18a), Ca2+/Calmodulin-dependent protein kinase I (CAMKI), creatine transporter (CRTR), CDM protein (CDM), adrenoleukodystrophy protein >
7058	19749	32812	2.28	9.7E-01	U26716.1	NT	Drosophila melanogaster sodium channel protein (para) gene, exons 9,10,11,12 and optional segments b, c, d and e, partial cds
8401	21094	34230	1.68	9.7E-01	AF149112.1	NT	Triticum aestivum stripe rust resistance protein Yr10 (Yr10) gene, complete cds
8407	21100	34236	1.3	9.7E-01	M90544.1	NT	Salmonella typhimurium adenine-methyltransferase (mod) and restriction endonuclease (res)
11123	23792		3.64	9.7E-01	BF511209.1	EST_HUMAN	UI-H-BI4-aol-e-07-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone IMAGE:3085140 3'
4425	17161	29791	1.5	9.6E-01	AW799674.1	EST_HUMAN	PM2-UM0053-240300-005-f12 UM0053 Homo sapiens cDNA
5667	18462	31376	3.77	9.6E-01	Z70556.1	NT	Parvovirus B19 DNA, patient C, genome position 2448-2994
5667	18462	31377	3.77	9.6E-01	Z70556.1	NT	Parvovirus B19 DNA, patient C, genome position 2448-2994
6648	19410	32424	0.61	9.6E-01	Z97341.2	NT	Arabidopsis thaliana DNA chromosome 4, ESSA I FCA contig fragment No. 6
8291	20985		2.33	9.6E-01	X95275.1	NT	P.falciparum complete gene map of plastid-like DNA (IR-A)
8750	21442	34589	0.59	9.6E-01	L81138.1	NT	Rattus norvegious (strain R21) Rps2r gene, complete cds
11503	24104	37416	3.47	9.6E-01	AV752605.1	EST_HUMAN	AV752605 NPD Homo sapiens cDNA clone NPDBAG06 5'
11503	24104	37417	3.47	9.6E-01	AV752605.1	EST_HUMAN	AV752605 NPD Homo sapiens cDNA clone NPDBAG06 5'
11952	24505		1.92	9.6E-01	11421722	NT	Homo sapiens centrosomal protein 2 (CEP2), mRNA
12558	25301	30710	2.18	9.6E-01	U91423.1	NT	Sphyrna tiburo NADH dehydrogenase subunit 2 (NADH2) gene, mitochondrial gene encoding mitochondrial protein, partial cds
2480	15198	27938	1.05	9.5E-01	7705591	NT	Homo sapiens CGI-125 protein (LOC51003), mRNA
2673	15382	28122	0.97	9.5E-01	Q02934	SWISSPROT	ENDOGLUCANASE I PRECURSOR (EGI) (ENDO-1,4-BETA-GLUCANASE) (CELLULASE I)
3762	16514	29150	2.04	9.5E-01	BE902340.1	EST_HUMAN	601675639F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3958473 5'
3762	16514	29151	2.04	9.5E-01	BE902340.1	EST_HUMAN	601675639F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3958473 5'
8899	21590	34730	0.69	9.5E-01	AI190162.1	EST_HUMAN	qd57d07.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1733581 3'
9003	21693	34843	1.05	9.5E-01	AW861102.1	EST_HUMAN	RC1-CT0295-241199-011-b02 CT0295 Homo sapiens cDNA

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11209	23872	37159	1.68	9.5E-01	BF218771.1	EST_HUMAN	601885163F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4103630 5'
11429	23196	36427	2.42	9.5E-01	AW293799.1	EST_HUMAN	UI-H-BI2-ahp-f-03-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone IMAGE:2727677 3'
11795	24385	37718	1.55	9.5E-01	T67204.1	EST_HUMAN	ya53d04.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:66631 3'
3196	15959		3.33	9.4E-01	AF165990.1	NT	Bartonella clarridgeiae RNA polymerase beta subunit (rpoB) gene, partial cds
3212	15975		2.06	9.4E-01	AF080595.1	NT	Pimpinella brachycarpa zinc finger protein (ZFP1) mRNA, complete cds
8764	21456	34606	0.67	9.4E-01	M90724.1	NT	Human Fc-gamma-receptorIIA (FCGR2A) gene, exon 4
12202	24670		1.92	9.4E-01	BE781251.1	EST_HUMAN	601466703F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3869929 5'
12557	25219		1.79	9.4E-01	11419857	NT	Homo sapiens epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog) (EGFR), mRNA
1726	14468		1.05	9.3E-01	AF242382.1	NT	Homo sapiens phytanoyl-CoA hydroxylase (PHYH) gene, exon 5
2640	15351	28095	1.36	9.3E-01	BE071172.1	EST_HUMAN	RC5-BT0503-271199-011-B01 BT0503 Homo sapiens cDNA
4015	16761	29388	0.88	9.3E-01	M20219.1	NT	Bovine papillomavirus type 2, complete genome
4015	16761	29389	0.88	9.3E-01	M20219.1	NT	Bovine papillomavirus type 2, complete genome
5505	18303	31204	1.56	9.3E-01	AF213884.1	NT	Homo sapiens nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1) gene, complete cds
5592	18388	31298	3.99	9.3E-01	L36189.1	NT	Spodoptera frugiperda methylenetetrahydrofolate dehydrogenase mRNA, complete cds
7966	20661	33785	1.65	9.3E-01	AA847040.1	EST_HUMAN	oe09b03.s1 NCI_CGAP_Ov2 Homo sapiens cDNA clone IMAGE:1385357
8713	21405		1.04	9.3E-01	AF061981.1	NT	Xenopus laevis CCCH zinc finger protein C3H-2 (C3H-2) mRNA, complete cds
8835	21527	34673	0.95	9.3E-01	AL161534.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 34
12681	24970		3.12	9.3E-01	AF271207.1	NT	Aedes triseriatus putative large subunit ribosomal protein rpL34 mRNA, complete cds
12802	25049		1.48	9.3E-01	U82671.2	NT	Homo sapiens chromosome Xq28 melanoma antigen family A2a (MAGEA2A), melanoma antigen family A12 (MAGEA12), melanoma antigen family A2b (MAGEA2B), melanoma antigen family A3 (MAGEA3), caltractin (CALT), NAD(P)H dehydrogenase-like protein (NSDHL), and LI>
3233	15995	28648	2.93	9.2E-01	BE622702.1	EST_HUMAN	601441338T1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3916184 3'
4822	17553		0.97	9.2E-01	BF129973.1	EST_HUMAN	601817814F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4041363 5'
5631	18426		1.15	9.2E-01	7106410	NT	Mus musculus solute carrier family 30 (zinc transporter), member 4 (Slc30a4), mRNA
5898	18683	31631	7.36	9.2E-01	BF037586.1	EST_HUMAN	601461153F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3864661 5'
6537	19302	32306	0.61	9.2E-01	M64703.1	NT	N.crassa valyl-tRNA synthetase (cyt-20/un-3) gene
9560	22213	35399	0.92	9.2E-01	AL161565.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 65
9648	22300	35496	1.07	9.2E-01	6671677	NT	Mus musculus carbonic anhydrase 4 (Car4), mRNA
10165	22813	36031	3.16	9.2E-01	11430963	NT	Homo sapiens lysosomal apyrase-like protein 1 (LALP1), mRNA
10315	22962	36178	1.9	9.2E-01	BF593251.1	EST_HUMAN	7c58e06.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3578219 3' similar to SW:NU5M_TRYBB P04540 NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 ;
10543	23239	36473	1.63	9.2E-01	BE563811.1	EST_HUMAN	601334943F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3688714 5'

Page 31 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11722	24316	37639	1.79	9.2E-01	BF132402.1	EST_HUMAN	601820312F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4052018 5'
1621	14368	27057	1.88	9.1E-01	T98675.1	EST_HUMAN	ye52f01.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:121369 3' similar to contains Alu repetitive element;
2120	14851		2.76	9.1E-01	8923056	NT	Homo sapiens hypothetical protein FLJ20048 (FLJ20048), mRNA
3200	15963	28614	1.15	9.1E-01	T26418.1	EST_HUMAN	AB200G8R Infant brain, LLNL array of Dr. M. Soares 1NIB Homo sapiens cDNA clone LLAB200G8 5'
3200	15963	28615	1.15	9.1E-01	T26418.1	EST_HUMAN	AB200G8R Infant brain, LLNL array of Dr. M. Soares 1NIB Homo sapiens cDNA clone LLAB200G8 5'
6075	18854	31821	1.28	9.1E-01	L36033.1	NT	Human pre-B cell stimulating factor homologue (SDF1b) mRNA, complete cds
6413	19181	32180	3.53	9.1E-01	Q61704	SWISSPROT	INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN H3 PRECURSOR (ITI HEAVY CHAIN H3)
7475	20148	33241	17.62	9.1E-01	AA806623.1	EST_HUMAN	ob71g08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336862 3'
7637	20302	33410	2.34	9.1E-01	U72995.1	NT	Rattus norvegicus Rab3 GDP/GTP exchange protein mRNA, complete cds
10075	22723	35940	0.45	9.1E-01	P38432	SWISSPROT	P80-COILIN
12291	25294		27.98	9.1E-01	AF050113.1	NT	Homo sapiens uncoupling protein-3 (UCP3) gene, complete cds
4348	17085	29714	2.08	9.0E-01	AF099810.1	NT	Homo sapiens neurexin III-alpha gene, partial cds
7291	19974	33052	0.72	9.0E-01	L42547.1	NT	Danio rerio LIM class homeodomain protein (lim5) mRNA, complete cds
7321	20004		1.18	9.0E-01	D38621.1	NT	Xenopus laevis gene for aldolase, complete cds
9249	21928	35100	0.49	9.0E-01	AF086761.1	NT	Danio rerio semaphorin Z1a mRNA, complete cds
5610	18406	31318	2.68	8.9E-01	AF026198.1	NT	Fugu rubripes neural cell adhesion molecule L1 homolog (L1-CAM) gene, complete cds; putative protein 1 (PUT1) gene, partial cds; mitosis-specific chromosome segregation protein SMC1 homolog (SMC1) gene, complete cds; and calcium channel alpha-1 subunit>
6154	18931		1.38	8.9E-01	X60986.1	NT	Rabbit MHC fragment RLA-DF DNA
8325	21018	34154	0.71	8.9E-01	AF259667.1	NT	Oithona nana cytochrome-c oxidase subunit I (coxI) gene, partial cds; mitochondrial gene for mitochondrial product
11787	24377	37707	2.51	8.9E-01	AE003944.1	NT	Xylella fastidiosa, section 90 of 229 of the complete genome
12138	24627		2.86	8.9E-01	AE002186.2	NT	Chlamydophila pneumoniae AR39, section 21 of 94 of the complete genome
12762	25343		2.51	8.9E-01	AI150836.1	EST_HUMAN	qb64d08.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:1704879 3'
4505	17240	29873	3.82	8.8E-01	O26350	SWISSPROT	PUTATIVE F420-DEPENDENT NADP REDUCTASE
5289	18094	30755	0.67	8.8E-01	AF310617.1	NT	Pseudorabies virus Ea glycoprotein M gene, complete cds
10131	22779	35992	0.83	8.8E-01	7656978	NT	Homo sapiens cell death-inducing DFFA-like effector B (CIDEB), mRNA
11018	23690	36953	4.96	8.8E-01	Z28337.1	NT	M.aeruginosa (HUB 5-2-4) DNA from plasmid PMA1
11968	25382		1.8	8.8E-01	D90911.1	NT	Synechocystis sp. PCC6803 complete genome, 13/27, 1576593-1719643
452	13238	25877	1.54	8.7E-01	AF106953.2	NT	Homo sapiens SOS1 (SOS1) gene, partial cds
2401	15122	27859	1.07	8.7E-01	5901893	NT	Homo sapiens AT-binding transcription factor 1 (ATBF1), mRNA

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2877	15644	28286	5.05	8.7E-01	AA595863.1	EST_HUMAN	nn05f11.s1 NCI_CGAP_Pr4.1 Homo sapiens cDNA clone IMAGE:1076877
4946	17673		3.17	8.7E-01	AF121970.1	NT	Pseudomonas aeruginosa topoisomerase (top), putative transcriptional regulatory protein OhbR (ohbR), ortho-halobenzoate 1,2-dioxygenase beta-ISP protein OhbA (ohbA), OhbC (ohbC), ortho-halobenzoate 1,2-dioxygenase alpha-ISP protein OhbB (ohbB), and put>
5102	17820		0.97	8.7E-01	AJ288085.1	NT	Homo sapiens partial LGALS9 gene for galectin-9, exon 3
7939	20634	33761	0.62	8.7E-01	AW897335.1	EST_HUMAN	RC4-NN0057-120500-013-c07 NN0057 Homo sapiens cDNA
8828	21520	34665	0.69	8.7E-01	AI239456.1	EST_HUMAN	qh36e06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1846786 3'
8828	21520	34666	0.69	8.7E-01	AI239456.1	EST_HUMAN	qh36e06.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1846786 3'
9638	22290	35483	1.57	8.7E-01	AE004963.1	NT	Pseudomonas aeruginosa PA01, section 524 of 529 of the complete genome
10202	22850	36065	0.61	8.7E-01	BF570169.1	EST_HUMAN	602185541T1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4309906 3'
10202	22850	36066	0.61	8.7E-01	BF570169.1	EST_HUMAN	602185541T1 NIH_MGC_45 Homo sapiens cDNA clone IMAGE:4309906 3'
10735	23422	36685	5.25	8.7E-01	BF363970.1	EST_HUMAN	QV0-NN1021-100800-337-c03 NN1021 Homo sapiens cDNA
11739	24332	37657	5.47	8.7E-01	BF107694.1	EST_HUMAN	601823684R1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4043564 3'
11739	24332	37658	5.47	8.7E-01	BF107694.1	EST_HUMAN	601823684R1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4043564 3'
462	13247		1.75	8.6E-01	X17012.1	NT	Rat IGFII gene for insulin-like growth factor II
838	13608	26279	3.45	8.6E-01	W69089.1	EST_HUMAN	zd44e03.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:343516 5'
2268	14994	27733	0.96	8.6E-01	4503210	NT	Homo sapiens cytochrome P450, subfamily XXVIIA (steroid 27-hydroxylase, cerebrotendinous xanthomatosis), polypeptide 1 (CYP27A1b) mRNA
3608	16361	29003	0.85	8.6E-01	AL161565.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 65
3782	16534	29172	1.55	8.6E-01	U49724.1	NT	Drosophila melanogaster merlin (Dmerlin) mRNA, complete cds
5808	18597	31524	10.86	8.6E-01	X60547.1	NT	Chicken lipoprotein lipase gene
5808	18597	31525	10.86	8.6E-01	X60547.1	NT	Chicken lipoprotein lipase gene
6609	19372	32385	2.06	8.6E-01	AF143732.1	NT	Grus canadensis recombination activating protein 1 (RAG-1) gene, partial cds
6609	19372	32386	2.06	8.6E-01	AF143732.1	NT	Grus canadensis recombination activating protein 1 (RAG-1) gene, partial cds
7427	20104		0.78	8.6E-01	AE000591.1	NT	Helicobacter pylori 26695 section 69 of 134 of the complete genome
7828	20523		1.12	8.6E-01	AP001518.1	NT	Bacillus halodurans genomic DNA, section 12/14
7941	20636	33763	0.55	8.6E-01	AF077837.1	NT	Drosophila melanogaster collapsin response mediator protein (CRMP) mRNA, complete cds
9585	22238		0.46	8.6E-01	AE000979.1	NT	Archaeoglobus fulgidus section 128 of 172 of the complete genome
12518	25144		1.35	8.6E-01	AL112162.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
6626	19388	32401	0.95	8.5E-01	AF165214.1	NT	Bacteriophage D3, complete genome
7425	20102	33189	2.51	8.5E-01	BE542612.1	EST_HUMAN	601067107F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3453505 5'
8317	21010	34147	0.78	8.5E-01	P06601	SWISSPROT	SEGMENTATION PROTEIN PAIRED
8317	21010	34148	0.78	8.5E-01	P06601	SWISSPROT	SEGMENTATION PROTEIN PAIRED
8402	21095	34231	0.67	8.5E-01	AJ243213.1	NT	Homo sapiens partial 5-HT4 receptor gene, exons 2 to 5

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10248	22896	36105	1.17	8.5E-01	AB006799.1	NT	Cyanidium caldarium gene for SigC, complete cds
10248	22896	36106	1.17	8.5E-01	AB006799.1	NT	Cyanidium caldarium gene for SigC, complete cds
12278	25296		2.24	8.5E-01	11418543	NT	Homo sapiens human immunodeficiency virus type I enhancer-binding protein 1 (HIVEP1), mRNA
4702	17436	30067	0.73	8.4E-01	AF083975.2	NT	Fowl adenovirus 8, complete genome
5406	25068	30910	2.28	8.4E-01	L78726.1	NT	Human fibroblast growth factor receptor 3 (FGFR3) gene, intron 7
5406	25068	30911	2.28	8.4E-01	L78726.1	NT	Human fibroblast growth factor receptor 3 (FGFR3) gene, intron 7
7708	20372	33485	0.63	8.4E-01	AF051142.1	NT	Mamestra brassicae pheromone binding protein 2 precursor (PBP2) mRNA, complete cds
9858	22508		2.68	8.4E-01	AJ248287.1	NT	Pyrococcus abyssi complete genome; segment 5/6
724	13498	26151	2.8	8.3E-01	M93437.1	NT	Thermus thermophilus cytochrome c-552 (cycA) and CycB (cycB) genes, complete cds
3091	15856	28497	2.99	8.3E-01	AL161506.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 18
3790	16542	29177	0.79	8.3E-01	AB010879.1	NT	Nicotiana tabacum mRNA for chloroplast ribosomal protein L10, complete cds
3993	16741	29375	3.35	8.3E-01	Y19177.1	NT	Streptomyces antibioticus polyketide biosynthetic gene cluster
5187	17995	30511	2.41	8.3E-01	AL161540.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 40
9568	22221		4.51	8.3E-01	AI791952.1	EST_HUMAN	nn01f12.y5 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:1076495 5' similar to contains THR.t1 THR repetitive element ;
10010	22658	35872	1.27	8.3E-01	AF098070.1	NT	Drosophila melanogaster Lis1 homolog mRNA, complete cds
10118	22766	35978	3.46	8.3E-01	AF108133.1	NT	Mus musculus neuro-d4 gene, exons 3 through 12 and partial cds
10572	23267	36505	3.35	8.3E-01	AE000903.1	NT	Methanobacterium thermoautotrophicum from bases 1270510 to 1283409 (section 109 of 148) of the complete genome
10590	23284		2.03	8.3E-01	7212472	NT	Phytophthora infestans mitochondrion, complete genome
11274	23935	37227	2	8.3E-01	AF020503.1	NT	Homo sapiens FRA3B common fragile region, diadenosine triphosphate hydrolase (FHIT) gene, exon 5
2045	14778	27506	2.3	8.2E-01	AB000489.1	NT	Rattus norvegicus mRNA for RPHO-1, complete cds
2083	14815		1.31	8.2E-01	AF145589.1	NT	Mus musculus trophinin (Tnn) gene, complete cds
2686	15395		1.06	8.2E-01	AW376990.1	EST_HUMAN	IL3-CT0219-161199-031-C08 CT0219 Homo sapiens cDNA
6676	19593	32631	0.76	8.2E-01	AJ010142.1	NT	Amanita muscaria mRNA for SCIII25 protein
6797	19541	32569	3.49	8.2E-01	AW379433.1	EST_HUMAN	CM4-HT0243-081199-037-e01 HT0243 Homo sapiens cDNA
7169	25106	32926	4.74	8.2E-01	Z12126.1	NT	S.cerevisiae MET, LEU4, and POL1 genes encoding MET4 protein, alpha-isopropylmalate (alpha-IPM) synthetase (partial), and DNA polymerase alpha (partial)
8343	21036	34173	0.59	8.2E-01	BE263145.1	EST_HUMAN	601144885F2 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3160412 5'
9924	22572	35770	0.65	8.2E-01	AB014530.1	NT	Homo sapiens mRNA for KIAA0630 protein, partial cds
9959	22607	35812	1.37	8.2E-01	AF052659.1	NT	Homo sapiens thioredoxin-related protein mRNA, complete cds
10123	22771	35985	0.59	8.2E-01	AF223888.1	NT	Oncorhynchus tshawytscha isolate T-20 somatolactin precursor gene, exon 1
10123	22771	35986	0.59	8.2E-01	AF223888.1	NT	Oncorhynchus tshawytscha isolate T-20 somatolactin precursor gene, exon 1

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10283	22931	36145	3.65	8.2E-01	Q9JI70	SWISSPROT	MCKUSICK-KAUFMAN/BARDET-BIEDL SYNDROMES PUTATIVE CHAPERONIN
10283	22931	36146	3.65	8.2E-01	Q9JI70	SWISSPROT	MCKUSICK-KAUFMAN/BARDET-BIEDL SYNDROMES PUTATIVE CHAPERONIN
11641	24238	37562	4.65	8.2E-01	L10127.1	NT	Molluscum contagiosum virus type 1 ORF1 and ORF2 DNA
11735	24328	37652	6.38	8.2E-01	P10383	SWISSPROT	OVARIAN TUMOR LOCUS PROTEIN
11740	24333	37659	6.02	8.2E-01	H87398.1	EST_HUMAN	yw14d02.r1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:252195 5' similar to gb:M36072 60S RIBOSOMAL PROTEIN L7A (HUMAN);
12298	24723	31054	2.37	8.2E-01	AJ001261.1	NT	Mus musculus mRNA for NIPSNAP2 protein
2762	15467		1.08	8.1E-01	AF191839.1	NT	Mus musculus TANK binding kinase TBK1 (Tbk1) mRNA, complete cds
3451	16207	28857	3.08	8.1E-01	AF055066.1	NT	Homo sapiens MHC class 1 region
3451	16207	28858	3.08	8.1E-01	AF055066.1	NT	Homo sapiens MHC class 1 region
4863	17592		0.74	8.1E-01	AF202634.1	NT	Drosophila melanogaster Na/K-ATPase beta subunit isoform 4 (JYbeta2) mRNA, complete cds
6223	18997	31973	0.64	8.1E-01	U16790.1	NT	Mus musculus putative collagen alpha-2 (XI) chain (COL11A2) gene, partial cds
6526	19292	32295	2.66	8.1E-01	Q13491	SWISSPROT	NEURONAL MEMBRANE GLYCOPROTEIN M6-B
6526	19292	32296	2.66	8.1E-01	Q13491	SWISSPROT	NEURONAL MEMBRANE GLYCOPROTEIN M6-B
7229	19914	32987	0.78	8.1E-01	AB007877.1	NT	Homo sapiens KIAA0417 mRNA, complete cds
7412	20089	33173	0.65	8.1E-01	O47477	SWISSPROT	CYTOCHROME B
7811	20506	33628	0.75	8.1E-01	AF022713.2	NT	Drosophila melanogaster putative inorganic phosphate cotransporter (Picot) gene, partial cds; putative sodium channel (Nach) and putative amylase-related protein (Amyrel) genes, complete cds; and putative serine-enriched protein (gprs) gene, partial cd>
7811	20506	33629	0.75	8.1E-01	AF022713.2	NT	Drosophila melanogaster putative inorganic phosphate cotransporter (Picot) gene, partial cds; putative sodium channel (Nach) and putative amylase-related protein (Amyrel) genes, complete cds; and putative serine-enriched protein (gprs) gene, partial cd>
8507	21199	34344	0.93	8.1E-01	AP001517.1	NT	Bacillus halodurans genomic DNA, section 11/14
8507	21199	34345	0.93	8.1E-01	AP001517.1	NT	Bacillus halodurans genomic DNA, section 11/14
8668	21360	34507	1.08	8.1E-01	AW242647.1	EST_HUMAN	xn01h03.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2692469 3' similar to SW:LYAR_MOUSE Q08288 CELL GROWTH REGULATING NUCLEOLAR PROTEIN. ;contains MER22.b1 PTR5,repetitive element ;
10025	22673	35888	0.7	8.1E-01	P06425	SWISSPROT	PROBABLE E4 PROTEIN
10311	22958	36174	0.5	8.1E-01	N84541.1	EST_HUMAN	KK9872F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA clone KK9872 5' similar to EST(CLONE C-0PE11)
11464	24067	37374	2.63	8.1E-01	BE938558.1	EST_HUMAN	RC0-TN0080-220800-025-d10 TN0080 Homo sapiens cDNA
11464	24067	37375	2.63	8.1E-01	BE938558.1	EST_HUMAN	RC0-TN0080-220800-025-d10 TN0080 Homo sapiens cDNA
12022	24550	31110	1.57	8.1E-01	AE001711.1	NT	Thermotoga maritima section 23 of 136 of the complete genome
172	12985		3.49	8.0E-01	AJ271510.1	NT	Staphylococcus aureus partial pta gene for phosphate actyltransferase allele 15

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
282	13089	25730	13.81	8.0E-01	AJ132772.1	NT	Bos taurus futb and rtif genes
1595	14341	27031	1.12	8.0E-01	8394087	NT	Rattus norvegicus protease (prosome, macropain) 28 subunit, alpha (Psme1), mRNA
2029	14764		1.91	8.0E-01	BF530962.1	EST_HUMAN	602072473F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4215091 5'
3075	15841	28484	1.2	8.0E-01	AF127897.1	NT	Saimiri boliviensis olfactory receptor (SBO27) gene, partial cds
3307	16067	28716	1.35	8.0E-01	AB006193.1	NT	Mus musculus gene for oviductal glycoprotein, complete cds
3690	16443		1.52	8.0E-01	AL162758.2	NT	Neisseria meningitidis serogroup A strain Z2491 complete genome; segment 7/7
4496	17232	29862	6.05	8.0E-01	X83739.2	NT	G.gallus mRNA for nicotinic acetylcholine receptor (nAChR) beta 3 subunit
7889	20584		2.25	8.0E-01	AW901489.1	EST_HUMAN	RC0-NN1012-270300-021-h06 NN1012 Homo sapiens cDNA
8423	21116	34254	0.98	8.0E-01	Y11095.1	NT	Rice stripe virus RNA 3
10876	23556	36803	2.78	8.0E-01	Q92793	SWISSPROT	CREB-BINDING PROTEIN
441	13227	25870	1.16	7.9E-01	D11476.1	NT	Lymantria dispar nuclear polyhedrosis virus gene for DNA polymerase, complete cds
698	13473		1.14	7.9E-01	AE002130.1	NT	Ureaplasma urealyticum section 31 of 59 of the complete genome
1600	14346		22.69	7.9E-01	AB040885.1	NT	Homo sapiens mRNA for KIAA1452 protein, partial cds
1652	14398		1.2	7.9E-01	U32739.1	NT	Haemophilus influenzae Rd section 54 of 163 of the complete genome
2259	14986	27726	5.66	7.9E-01	AB004816.1	NT	Oryctolagus cuniculus mRNA for mitsugumin29, complete cds
2260	14987	27727	2.4	7.9E-01	AF130459.1	NT	Danio rerio Trp4-associated protein Tap1A (tap1A) mRNA, complete cds
3506	16262	28916	3.01	7.9E-01	AF228664.1	NT	Gallus gallus SOX8 transcription factor (SOX8) mRNA, complete cds
4268	17008		0.85	7.9E-01	BE263612.1	EST_HUMAN	601192033F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3535785 5'
4572	17307	29935	1.13	7.9E-01	6753745	NT	Mus musculus embigin (Emb), mRNA
4572	17307	29936	1.13	7.9E-01	6753745	NT	Mus musculus embigin (Emb), mRNA
6252	19026	32000	0.67	7.9E-01	D38145.1	NT	Human mRNA for prostacyclin synthase, complete cds
8008	20703	33831	2.79	7.9E-01	X90996.1	NT	P.sativum GR gene
9447	22124	35304	4.04	7.9E-01	U01912.1	NT	Giardia lamblia variant-specific surface protein G3M-B (vspG3M-B) mRNA, partial cds
9949	22597	35801	4.47	7.9E-01	P19719	SWISSPROT	SMALL HYDROPHOBIC PROTEIN
9991	22639	35849	0.91	7.9E-01	AV700860.1	EST_HUMAN	AV700860 GKC Homo sapiens cDNA clone GKCDRE12 3'
10408	23054	36271	1.94	7.9E-01	AB000631.1	NT	Streptococcus mutans DNA for sigma 42 protein, dTDP-4-keto-L-rhamnose reductase, complete cds
10516	23162	36389	0.52	7.9E-01	P15305	SWISSPROT	DYNEIN HEAVY CHAIN (DYHC)
10929	23609		2.74	7.9E-01	7662471	NT	Homo sapiens KIAA1072 protein (KIAA1072), mRNA
11173	23840	37123	2.02	7.9E-01	P19022	SWISSPROT	NEURAL-CADHERIN PRECURSOR (N-CADHERIN)
856	13625		2.24	7.8E-01	Z43785.1	EST_HUMAN	HSC1KH041 normalized infant brain cDNA Homo sapiens cDNA clone c-1kh04
2273	14999	27737	7.47	7.8E-01	AW959567.1	EST_HUMAN	EST371637 MAGE resequences, MAGF Homo sapiens cDNA
4653	17387	30020	0.73	7.8E-01	U87305.1	NT	Rattus norvegicus transmembrane receptor Unc5H1 mRNA, complete cds
5978	18760	31724	2.28	7.8E-01	AF115856.1	NT	Sphenodon punctatus alpha enolase mRNA, partial cds

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6124	18902	31871	0.88	7.8E-01	P05231	SWISSPROT	INTERLEUKIN-6 PRECURSOR (IL-6) (B-CELL STIMULATORY FACTOR 2) (BSF-2) (INTERFERON BETA-2) (HYBRIDOMA GROWTH FACTOR)
6371	19140	32136	0.63	7.8E-01	AL445066.1	NT	Thermoplasma acidophilum complete genome; segment 4/5
8389	21082	34216	1.02	7.8E-01	BF108927.1	EST_HUMAN	7154d05.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:3525176 3'
9133	21821	34987	1.3	7.8E-01	Y10159.1	NT	D.discoideum racGAP gene
9231	21910	35083	0.51	7.8E-01	4826873	NT	Homo sapiens nucleoporin 214kD (CAIN) (NUP214), mRNA
10024	22672		0.97	7.8E-01	Q25452	SWISSPROT	MUSCLE CALCIUM CHANNEL ALPHA-1 SUBUNIT (MDL-ALPHA1)
12271	25275		2.5	7.8E-01	L29260.1	NT	Arabidopsis thaliana 1-amino-1-cyclopropanecarboxylate synthase (ACS5) gene, complete cds
139	12954	25596	7.61	7.7E-01	AF184345.1	NT	Lycopersicon hirsutum ADP-glucose pyrophosphorylase large subunit (AGP-L1) mRNA, complete cds
709	13483		2.26	7.7E-01	AF050157.1	NT	Mus musculus major histocompatibility locus class II region: major histocompatibility protein class II alpha chain (IAalpha) and major histocompatibility protein class II beta chain (IEbeta) genes, complete cds; butyrophilin-like (NG9), butyrophilin-li>
2717	15424	28163	2.21	7.7E-01	O33915	SWISSPROT	CITRATE SYNTHASE
3351	16111		0.84	7.7E-01	8393408	NT	Homo sapiens UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GalNAc-T7) (GALNAC-T7), mRNA
3586	16340	28985	3.98	7.7E-01	AF118085.1	NT	Homo sapiens PRO1975 mRNA, complete cds
4365	17103	29738	3.38	7.7E-01	AF199488.1	NT	Coturnix coturnix japonica sub-species japonica beta-actin mRNA, partial cds
4365	17103	29739	3.38	7.7E-01	AF199488.1	NT	Coturnix coturnix japonica sub-species japonica beta-actin mRNA, partial cds
5473	18272	31165	1.45	7.7E-01	P16553	SWISSPROT	RAFFINOSE INVERTASE (INVERTASE)
5473	18272	31166	1.45	7.7E-01	P16553	SWISSPROT	RAFFINOSE INVERTASE (INVERTASE)
5866	18653	31594	0.85	7.7E-01	R08600.1	EST_HUMAN	yf24b02.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:127755 3'
9744	22395	35600	0.51	7.7E-01	AB021134.1	NT	Daphnia magna hemoglobin gene cluster (dhb3, dhb1 and dhb2 genes), complete cds
12161	24644		4.55	7.7E-01	11497621	NT	Archaeoglobus fulgidus, complete genome
6008	18789	31751	4.88	7.6E-01	AF059510.1	NT	Arabidopsis thaliana 3-methylcrotonyl-CoA carboxylase non-biotinylated subunit (MCCB) mRNA, complete cds
6008	18789	31752	4.88	7.6E-01	AF059510.1	NT	Arabidopsis thaliana 3-methylcrotonyl-CoA carboxylase non-biotinylated subunit (MCCB) mRNA, complete cds
6425	19193	32189	0.81	7.6E-01	P37938	SWISSPROT	MATING-TYPE PROTEIN A-ALPHA Z4
6751	17920	30555	0.94	7.6E-01	AI253399.1	EST_HUMAN	aq14b12.x1 Stanley Frontal NS pool 2 Homo sapiens cDNA clone IMAGE:2030879
6751	17920	30585	0.94	7.6E-01	AI253399.1	EST_HUMAN	aq14b12.x1 Stanley Frontal NS pool 2 Homo sapiens cDNA clone IMAGE:2030879
6951	19433	32449	0.88	7.6E-01	U72487.1	NT	Rattus norvegicus calcium-independent alpha-latrotoxin receptor mRNA, complete cds

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7964	20659	33784	1.38	7.6E-01	AF146793.2	NT	Mus musculus neuromedin U precursor (Nmu) gene, partial cds; tPhLP (Tphlp) gene, partial cds; CLOCK (Clock) gene, complete cds; PFT27 (Pft27) gene, complete cds; and H5AR (H5ar) gene, complete cds
8026	20721	33852	1.88	7.6E-01	6857752	NT	Mus musculus advillin (Advil-pending), mRNA
8026	20721	33853	1.88	7.6E-01	6857752	NT	Mus musculus advillin (Advil-pending), mRNA
8866	21557	34703	0.74	7.6E-01	6753577	NT	Mus musculus cytochrome P450, 2b9, phenobarbitol inducible, type a (Cyp2b9), mRNA
9179	21849	35015	5.03	7.6E-01	P30372	SWISSPROT	MUSCARINIC ACETYLCHOLINE RECEPTOR M2
9179	21849	35016	5.03	7.6E-01	P30372	SWISSPROT	MUSCARINIC ACETYLCHOLINE RECEPTOR M2
11330	24021	37325	2.68	7.6E-01	X86347.1	NT	H.aspersa mRNA for neurofilament NF70
11330	24021	37326	2.68	7.6E-01	X86347.1	NT	H.aspersa mRNA for neurofilament NF70
11711	24306		3.64	7.6E-01	AL161592.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 88
11931	24489		3.73	7.6E-01	AB020702.1	NT	Homo sapiens mRNA for KIAA0895 protein, partial cds
500	13284		1.44	7.5E-01	AL163301.2	NT	Homo sapiens chromosome 21 segment HS21C101
570	13351	25979	1.23	7.5E-01	AF020503.1	NT	Homo sapiens FRA3B common fragile region, diadenosine triphosphate hydrolase (FHIT) gene, exon 5
3354	16114	28769	0.95	7.5E-01	C14203.1	EST_HUMAN	C14203 Clontech human aorta polyA+ mRNA (#6572) Homo sapiens cDNA clone GEN-037E11 5'
7421	20098	33186	1.01	7.5E-01	AF052730.1	NT	Drosophila melanogaster tyrosine kinase receptor protein (eph) mRNA, complete cds
11177	23844	37130	1.5	7.5E-01	AB047819.1	NT	Homo sapiens GCMa/GCM1 gene for chorion-specific transcription factor GCMa, complete cds
.12228	24682		4.8	7.5E-01	AF163151.2	NT	Homo sapiens dentin sialophosphoprotein precursor (DSPP) gene, complete cds
12742	25008	30975	1.46	7.5E-01	AE000823.1	NT	Methanobacterium thermoautotrophicum from bases 317350 to 328792 (section 29 of 148) of the complete genome
1108	13866	26522	1.78	7.4E-01	AI598146.1	EST_HUMAN	tn14b09.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2167577 3' similar to contains Alu repetitive element;contains element MIR repetitive element ;
2342	15065	27802	0.96	7.4E-01	AB011106.1	NT	Homo sapiens mRNA for KIAA0534 protein, partial cds
4276	17015	29642	4.73	7.4E-01	AL163246.2	NT	Homo sapiens chromosome 21 segment HS21C046
7743	20439	33562	1.23	7.4E-01	AL161551.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 51
7743	20439	33563	1.23	7.4E-01	AL161551.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 51
8531	21223	34365	0.83	7.4E-01	BF346266.1	EST_HUMAN	602018456F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4154340 5'
8613	21305		0.76	7.4E-01	U87960.1	NT	Rattus norvegicus leukocyte common antigen receptor (LAR) gene, trans-spliced alternative untranslated exon
8994	21684	34834	6.95	7.4E-01	BE747503.1	EST_HUMAN	601573026F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3834174 5'
9054	21743	34901	1.14	7.4E-01	AA187986.1	EST_HUMAN	zp87h01.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone IMAGE:625297 3' similar to SW:TCPQ_MOUSE P42932 T-COMPLEX PROTEIN 1, THETA SUBUNIT ;
10302	22949	36164	0.76	7.4E-01	11424933	NT	Homo sapiens NY-REN-45 antigen (LOC51133), mRNA

Page 38 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11665	24260	37582	1.65	7.4E-01	AB021490.2	NT	Oryzias latipes gene for membrane guanylyl cyclase OlGC1, complete cds
11665	24260	37583	1.65	7.4E-01	AB021490.2	NT	Oryzias latipes gene for membrane guanylyl cyclase OlGC1, complete cds
11900	24467		3.62	7.4E-01	6753217	NT	Mus musculus complement component 1 inhibitor (C1nh), mRNA
12008	24542		1.78	7.4E-01	AI472641.1	EST_HUMAN	ta13h01.x1 NCI_CGAP_Lym5 Homo sapiens cDNA clone IMAGE:2043985 3'
2999	15765	28413	0.8	7.3E-01	P09710	SWISSPROT	HYPOTHETICAL PROTEIN HKLF1 (IRL1) (TRL1)
4575	17310	29938	0.7	7.3E-01	AE001166.1	NT	Borrelia burgdorferi (section 52 of 70) of the complete genome
4652	17386	30019	4.37	7.3E-01	AF225421.1	NT	Homo sapiens HT017 mRNA, complete cds
5040	17759	30373	1.01	7.3E-01	O43103	SWISSPROT	FERRICHROME SIDEROPHORE PEPTIDE SYNTHETASE
6511	19276	32276	5.92	7.3E-01	L35772.1	NT	Mus musculus antigen (CD72) gene
6511	19276	32277	5.92	7.3E-01	L35772.1	NT	Mus musculus antigen (CD72) gene
6994	25103	32735	0.67	7.3E-01	AJ011418.1	NT	Lycopersicon esculentum mRNA for ubiquitin activating enzyme
7359	20040	33118	0.66	7.3E-01	Z14133.1	NT	D.melanogaster Chc mRNA for clathrin heavy chain
7445	20121	33210	7.84	7.3E-01	M26511.1	NT	V.alginolyticus sucrase (scrB) gene, complete cds
7445	20121	33211	7.84	7.3E-01	M26511.1	NT	V.alginolyticus sucrase (scrB) gene, complete cds
11407	24056	37361	3.83	7.3E-01	AA678019.1	EST_HUMAN	zi25b08.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:431799 3'
11407	24056	37362	3.83	7.3E-01	AA678019.1	EST_HUMAN	zi25b08.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:431799 3'
812	13583		3.89	7.2E-01	L29281.1	NT	Rattus norvegicus initiation factor-2 kinase (eIF-2a) mRNA, complete cds
1950	14685	27398	2.32	7.2E-01	X79140.1	NT	N.tabacum NeIF-4A13 mRNA
2463	15181	27920	1.27	7.2E-01	AB009605.1	NT	Gallus gallus gene for melanocortin 2-receptor, complete cds
3063	15829	28473	1.38	7.2E-01	AF198100.1	NT	Fowlpox virus, complete genome
3445	16201	28851	2.56	7.2E-01	AF065606.1	NT	Giardia intestinalis variant-specific surface protein (vsp417-6) gene, vsp417-6/A-I allele, complete cds
3601	16354	28994	1.06	7.2E-01	AB002307.1	NT	Human mRNA for KIAA0309 gene, partial cds
4040	16785		0.7	7.2E-01	AF108093.1	NT	Homo sapiens IA-2 gene, intron 18
4718	17450	30083	2.65	7.2E-01	D90314.1	NT	L.mesenteroides gene for sucrose phosphorylase (EC 2.4.1.7)
5075	17794	30410	0.74	7.2E-01	P33066	SWISSPROT	NUCLEOSIDE TRIPHOSPHATASE I (NUCLEOSIDE TRIPHOSPHATE PHOSPHOHYDROLASE I) (NPH I)
7112	19800	32864	0.88	7.2E-01	U69633.1	NT	Solanum tuberosum cold-stress inducible protein (C17) gene, complete cds
8353	21046	34183	1.11	7.2E-01	AF236061.1	NT	Oryctolagus cuniculus RING-finger binding protein mRNA, partial cds
8862	21553		0.46	7.2E-01	AV743773.1	EST_HUMAN	AV743773 CB Homo sapiens cDNA clone CBMAFD06 5'
10239	22887	36100	2.33	7.2E-01	BF670061.1	EST_HUMAN	602118381F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4275381 5'
10639	23330	36568	4.02	7.2E-01	U82623.1	NT	Rattus norvegicus cytocentrin mRNA, complete cds
11104	23774	37049	1.27	7.2E-01	S76838.1	NT	Dbs=Dbl guanine nucleotide exchange factor homolog [mice, 32D murine hemopoietic cell line, mRNA, 3923 nt]

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
12422	24798		2.9	7.2E-01	AP000063.1	NT	Aeropyrum pernix genomic DNA, section 6/7
676	13451	26094	12.73	7.1E-01	D21070.1	NT	Rana catesbeiana mRNA for bullfrog skeletal muscle calcium release channel (ryanodine receptor) alpha isoform(RyR1), complete cds
3059	15825	28470	11.76	7.1E-01	AJ270777.1	NT	Homo sapiens partial TCF-4 gene for T-cell transcription factor-4, exons 15-16
4184	16925	29555	3.18	7.1E-01	7305360	NT	Mus musculus otogelin (Otog), mRNA
4184	16925	29556	3.18	7.1E-01	7305360	NT	Mus musculus otogelin (Otog), mRNA
5858	18645	31585	1.63	7.1E-01	BF681034.1	EST_HUMAN	602155438F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4296344 5'
5858	18645	31586	1.63	7.1E-01	BF681034.1	EST_HUMAN	602155438F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4296344 5'
6850	19550	32580	7.68	7.1E-01	U36232.1	NT	Drosophila melanogaster 6-pyruvoyltetrahydropterin synthase (pr) gene, complete cds
8091	20785	33916	0.56	7.1E-01	H54244.1	EST_HUMAN	yq89d09.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:202961 3'
8635	21327	34469	0.93	7.1E-01	BE074185.1	EST_HUMAN	RC1-BT0567-301299-011-d09 BT0567 Homo sapiens cDNA
8635	21327	34470	0.93	7.1E-01	BE074185.1	EST_HUMAN	RC1-BT0567-301299-011-d09 BT0567 Homo sapiens cDNA
9755	22406	35613	1.43	7.1E-01	BE904405.1	EST_HUMAN	601496330F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3898495 5'
10309	22956	36172	1.22	7.1E-01	M12961.1	NT	Human T-cell receptor germline gamma-chain J2 gene
12211	25205		2.21	7.1E-01	AA421492.1	EST_HUMAN	zu05h11.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:731109 3'
1207	13958	26624	0.99	7.0E-01	AB014514.1	NT	Homo sapiens mRNA for KIAA0614 protein, partial cds
1207	13958	26625	0.99	7.0E-01	AB014514.1	NT	Homo sapiens mRNA for KIAA0614 protein, partial cds
2450	15169	27907	1.13	7.0E-01	N62412.1	EST_HUMAN	yz73e07.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone IMAGE:288708 3' similar to contains Alu repetitive element;
2450	15169	27908	1.13	7.0E-01	N62412.1	EST_HUMAN	yz73e07.s1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone IMAGE:288708 3' similar to contains Alu repetitive element;
4996	17719		1.78	7.0E-01	AL163301.2	NT	Homo sapiens chromosome 21 segment HS21C101
5862	18649		1.11	7.0E-01	AB021316.1	NT	Arabidopsis thaliana mRNA for chlorophyll b synthase, complete cds
6276	20970		11.76	7.0E-01	AE000253.1	NT	Escherichia coli K-12 MG1655 section 143 of 400 of the complete genome
9216	21895	35064	0.57	7.0E-01	U53868.1	NT	Clostridium acetobutylicum mannitol-specific phosphotransferase system (PTS) system, mtlA, mtlR, mtlF, and mtlD genes, complete cds
9216	21895	35065	0.57	7.0E-01	U53868.1	NT	Clostridium acetobutylicum mannitol-specific phosphotransferase system (PTS) system, mtlA, mtlR, mtlF, and mtlD genes, complete cds
10526	23172	36400	0.49	7.0E-01	U34662.1	NT	Danio rerio complement factor B mRNA, complete cds
11064	23734	37008	1.94	7.0E-01	AV763842.1	EST_HUMAN	AV763842 MDS Homo sapiens cDNA clone MDSCHE04 5'
11064	23734	37007	1.94	7.0E-01	AV763842.1	EST_HUMAN	AV763842 MDS Homo sapiens cDNA clone MDSCHE04 5'
949	13715	26380	11.02	6.9E-01	U69874.1	NT	Candida albicans squalene epoxidase (CAERG1) gene, complete cds and translational regulator gene, partial cds

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
949	13715	26381	11.02	6.9E-01	U69674.1	NT	Candida albicans squalene epoxidase (CAERG1) gene, complete cds and translational regulator gene, partial cds
1287	14037	26708	2.74	6.9E-01	AA593530.1	EST_HUMAN	nn28a09.s1 NCI_CGAP_Gas1 Homo sapiens cDNA clone IMAGE:1085176 3'
3213	15976	28627	1.97	6.9E-01	AE002271.2	NT	Chlamydia muridarum, section 3 of 85 of the complete genome
5694	18488	31409	0.91	6.9E-01	AB035662.1	NT	Branchiostoma belcheri BbNA3 mRNA for notochord actin, complete cds
5900	18685	31633	0.82	6.9E-01	Y18278.1	NT	Drosophila melanogaster mRNA for A-kinase anchor protein DAKAP550, partial
6277	19050	32027	1.36	6.9E-01	BE296188.1	EST_HUMAN	601177333F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3532328 5'
7697	20360	33474	0.65	6.9E-01	AF248863.1	NT	Strongylocentrotus purpuratus myosin V, complete cds
7879	20574	33700	2.96	6.9E-01	AL161573.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 69
7879	20574	33701	2.96	6.9E-01	AL161573.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 69
9069	21758		0.79	6.9E-01	AF118046.1	NT	Entamoeba dispar cation transporting ATPase (atpase) gene, partial cds
9594	22247	35431	0.59	6.9E-01	AF206319.1	NT	Musa acuminata pectate lyase 1 (PL1) mRNA, complete cds
9594	22247	35432	0.59	6.9E-01	AF206319.1	NT	Musa acuminata pectate lyase 1 (PL1) mRNA, complete cds
11223	23886	37172	2.38	6.9E-01	D89013.1	NT	Homo sapiens DAN gene, complete cds
11223	23886	37173	2.38	6.9E-01	D89013.1	NT	Homo sapiens DAN gene, complete cds
11878	25197		3.01	6.9E-01	Q99958	SWISSPROT	FORKHEAD BOX PROTEIN C2 (FORKHEAD-RELATED PROTEIN FKHL14) (MESENCHYME FORK HEAD PROTEIN 1) (MFH-1 PROTEIN) (TRANSCRIPTION FACTOR FKH-14)
937	13704	26369	1.05	6.8E-01	AF017784.1	NT	Giardia intestinalis carbamate kinase gene, complete cds
2680	15389		0.99	6.8E-01	D90917.1	NT	Synechocystis sp. PCC6803 complete genome, 27/27, 3418852-3573470
2832	14356	27045	1.49	6.8E-01	AA854475.1	EST_HUMAN	aj75a05.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone IMAGE:1402256 3' similar to gb:X56411_rna1 ALCOHOL DEHYDROGENASE CLASS II PI CHAIN (HUMAN);
4533	17268	29901	1.45	6.8E-01	J00762.1	NT	Rat(hooded) prolactin gene : exon iii and flanks
9538	22191	35375	1.45	6.8E-01	AB037766.1	NT	Homo sapiens mRNA for KIAA1345 protein, partial cds
11027	23699	36962	1.92	6.8E-01	AJ276675.1	NT	Stagonospora avenae bgl1 gene for beta-glucosidase, exons 1-4
11027	23699	36963	1.92	6.8E-01	AJ276675.1	NT	Stagonospora avenae bgl1 gene for beta-glucosidase, exons 1-4
11058	23728	37000	2.4	6.8E-01	AF038939.1	NT	Mus musculus zinc finger protein (Peg3) mRNA, complete cds
11058	23728	37001	2.4	6.8E-01	AF038939.1	NT	Mus musculus zinc finger protein (Peg3) mRNA, complete cds
11607	24205	37527	1.36	6.8E-01	AF110520.1	NT	Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial>
11607	24205	37528	1.36	6.8E-01	AF110520.1	NT	Mus musculus major histocompatibility complex region NG27, NG28, RPS28, NADH oxidoreductase, NG29, KIFC1, Fas-binding protein, BING1, tapasin, RalGDS-like, KE2, BING4, beta 1,3-galactosyl transferase, and RPS18 genes, complete cds; Sacm21 gene, partial>

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
291	13097	25739	44.11	6.7E-01	AF213884.1	NT	Homo sapiens nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1) gene, complete cds
330	13131	25766	21.34	6.7E-01	AF213884.1	NT	Homo sapiens nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (NFKB1) gene, complete cds
2143	14873	27606	1.73	6.7E-01	AA451884.1	EST_HUMAN	zx12g12.s1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:786310 3' similar to contains element TAR1 repetitive element ;
2163	15587	27628	2.51	6.7E-01	AF186073.1	NT	Drosophila melanogaster Mst85C gene, complete cds; NMDMC isoform (Nmdmc) gene, complete cds, alternatively spliced; and transcription factor (Relish) gene, complete cds, alternatively spliced
2994	15760	28408	3.41	6.7E-01	6678580	NT	Mus musculus Wiskott-Aldrich syndrome protein (Wasp), mRNA
4419	17155	29786	0.79	6.7E-01	X74421.1	NT	S.tuberosum mRNA for glucose-6-phosphate dehydrogenase
5422	18221	30932	0.94	6.7E-01	J04836.1	NT	M.barkeri ATPase alpha and beta subunit (atpA and atpB) genes, complete cds
5422	18221	30933	0.94	6.7E-01	J04836.1	NT	M.barkeri ATPase alpha and beta subunit (atpA and atpB) genes, complete cds
6231	19005	31981	1.18	6.7E-01	9635035	NT	Gallid herpesvirus 2, complete genome
6231	19005	31982	1.18	6.7E-01	9635035	NT	Gallid herpesvirus 2, complete genome
7215	19900		4.34	6.7E-01	AE004606.1	NT	Pseudomonas aeruginosa PA01, section 167 of 529 of the complete genome
7240	19925	33000	0.92	6.7E-01	AE001486.1	NT	Helicobacter pylori, strain J99 section 47 of 132 of the complete genome
10044	22692		0.68	6.7E-01	M34046.1	NT	Human placental protein 14 (PP14) gene, complete cds
10873	23553	36800	2.07	6.7E-01	BF354649.1	EST_HUMAN	CM3-HT0769-010600-197-c03 HT0769 Homo sapiens cDNA
11436	23203	36435	3.59	6.7E-01	O14357	SWISSPROT	N-ACETYLGLUCOSAMINYL-PHOSPHATIDYLINOSITOL BIOSYNTHETIC PROTEIN GPI1
11659	24255	37578	1.66	6.7E-01	AA342521.1	EST_HUMAN	EST48065 Fetal spleen Homo sapiens cDNA 3' end
2505	15222	27964	1.29	6.6E-01	AF075240.1	NT	Homo sapiens SLIT1 protein (SLIL2) mRNA, partial cds
2704	15411	28148	1.44	6.6E-01	AF199339.1	NT	Homo sapiens lens epithelium-derived growth factor gene, alternatively spliced, complete cds
3650	16403	29043	4.57	6.6E-01	Y07669.1	NT	C.albicans random DNA marker, 282bp
4089	16832		0.85	6.6E-01	U91328.1	NT	Human hereditary haemochromatosis region, histone 2A-like protein gene, hereditary haemochromatosis (HLA-H) gene, RoRet gene, and sodium phosphate transporter (NPT3) gene, complete cds
5125	17843	30461	1.13	6.6E-01	AL161572.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 68
6240	19014	31988	4.29	6.6E-01	6680577	NT	Mus musculus kinesin light chain 2 (Klc2), mRNA
7585	20253	33359	3.76	6.6E-01	AV660506.1	EST_HUMAN	AV660506 GLC Homo sapiens cDNA clone GLCGID04 3'
8464	21156	34299	0.52	6.6E-01	AV704700.1	EST_HUMAN	AV704700 ADB Homo sapiens cDNA clone ADBCAF11 5'
9564	22217		2	6.6E-01	AL163278.2	NT	Homo sapiens chromosome 21 segment HS21C078
12470	24836	31033	1.48	6.6E-01	AE004382.1	NT	Vibrio cholerae chromosome II, section 39 of 93 of the complete chromosome
610	13388	26019	18.23	6.5E-01	M75140.1	NT	H.vulgaris Na,K-ATPase alpha subunit mRNA, complete cds

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
610	13388	26020	18.23	6.5E-01	M75140.1	NT	H.vulgaris Na,K-ATPase alpha subunit mRNA, complete cds
3426	16183	28833	4.25	6.5E-01	AB041225.1	NT	Mus musculus gene for Tob2, complete cds
4249	16990	29615	4.23	6.5E-01	AJ272265.1	NT	Homo sapiens SPP2 gene for secreted phosphoprotein 24 precursor, exons 1-8
4277	17016	29643	0.78	6.5E-01	AL161539.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 39
5003	17726	30329	2.6	6.5E-01	U28921.1	NT	Phaseolus vulgaris ATPase gamma subunit mRNA, nuclear gene encoding mitochondrial protein, partial cds
5357	25067	30843	1.77	6.5E-01	P18480	SWISSPROT	TRANSCRIPTION REGULATORY PROTEIN SNF5 (SWI/SNF COMPLEX COMPONENT SNF5) (TRANSCRIPTION FACTOR TYE4)
5627	18424	31337	0.62	6.5E-01	AL163249.2	NT	Homo sapiens chromosome 21 segment HS21C049
6625	19387	32400	1.5	6.5E-01	D88348.1	NT	Chicken mRNA for 115-kDa melanosomal matrix protein, complete cds
7566	20236	33340	0.84	6.5E-01	AI799882.1	EST_HUMAN	wc46a02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2321642 3'
9737	22388		0.8	6.5E-01	T78904.1	EST_HUMAN	yd21b04.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:108847 3'
10233	22881	36094	1.96	6.5E-01	AF119676.1	NT	Mus musculus small GTP-binding protein RAB25 (Rab25) gene, complete cds
10529	23226	36460	2.68	6.5E-01	H87583.1	EST_HUMAN	yw17f08.r1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:252515 5'
10585	23280	36518	3.5	6.5E-01	AA601287.1	EST_HUMAN	no15c07.s1 NCI_CGAP_Phe1 Homo sapiens cDNA clone IMAGE:1100748 3'
10690	23381		3.93	6.5E-01	AU138078.1	EST_HUMAN	AU138078 PLACE1 Homo sapiens cDNA clone PLACE1007810 5'
11599	24198	37518	2.42	6.5E-01	AF014115.1	NT	Plasmodium berghei cytochrome c oxidase subunit III, cytochrome c oxidase subunit I, and cytochrome b genes, mitochondrial genes encoding mitochondrial proteins, complete cds
12267	24710		2.07	6.5E-01	BE465050.1	EST_HUMAN	hv74a10.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3179130 3'
12504	25146		1.81	6.5E-01	Z74145.1	NT	S.cerevisiae chromosome IV reading frame ORF YDL097c
245	13054	25694	8.05	6.4E-01	U48848.1	NT	Drosophila melanogaster 8kd dynein light chain mRNA, complete cds
2593	15307	28043	1.16	6.4E-01	AF161184.1	NT	Pseudomonas fluorescens tryptophan halogenase (prnA) gene, complete cds
3449	16205	28855	2.16	6.4E-01	U48854.2	NT	Mus musculus dystroglycan 1 (DAG1) gene, exons 1 and 2 and complete cds
3842	16593	29230	1.06	6.4E-01	AB046827.1	NT	Homo sapiens mRNA for KIAA1607 protein, partial cds
8510	21202	34347	1.82	6.4E-01	AE001247.1	NT	Treponema pallidum section 63 of 87 of the complete genome
9989	22637	35848	8.6	6.4E-01	U82828.1	NT	Homo sapiens ataxia telangiectasia (ATM) gene, complete cds
10004	22652	35864	1.22	6.4E-01	BF670405.1	EST_HUMAN	602150289F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4291126 5'
12382	24777		5.99	6.4E-01	AV759212.1	EST_HUMAN	AV759212 MDS Homo sapiens cDNA clone MDSCGC09 5'
425	13211	25858	4.58	6.3E-01	P05228	SWISSPROT	HISTIDINE-RICH PROTEIN PRECURSOR (CLONE PFHRP-III)
522	13306	25938	2.25	6.3E-01	U32689.1	NT	Haemophilus influenzae Rd section 4 of 163 of the complete genome
2159	14889	27623	2.02	6.3E-01	U81136.1	NT	Shigella flexneri multi-antibiotic resistance locus
2583	15297	28035	3.51	6.3E-01	U75331.1	NT	Gallus gallus bone morphogenetic protein 1 (BMP1) mRNA, partial cds
2583	15297	28036	3.51	6.3E-01	U75331.1	NT	Gallus gallus bone morphogenetic protein 1 (BMP1) mRNA, partial cds

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5973	18755	31716	0.94	6.3E-01	BE093906.1	EST_HUMAN	PM0-BT0757-010500-002-a05 BT0757 Homo sapiens cDNA
6504	19269	32271	0.84	6.3E-01	L27798.1	NT	Streptococcus dysgalactiae (mag) gene, complete cds
6504	19269	32272	0.84	6.3E-01	L27798.1	NT	Streptococcus dysgalactiae (mag) gene, complete cds
8419	21112		3.44	6.3E-01	BE902044.1	EST_HUMAN	601676889F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959351 5'
8784	21476	34624	0.95	6.3E-01	S62927.1	NT	glycoprotein IIIa {Alu 1 and 3 fusion junction} [human, Genomic Mutant, 300 nt]
9120	21808	34975	0.8	6.3E-01	BF216984.1	EST_HUMAN	601884050F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4102596 5'
9320	21987	35159	2.45	6.3E-01	9627521	NT	Variola virus, complete genome
9320	21987	35160	2.45	6.3E-01	9627521	NT	Variola virus, complete genome
9838	22489		0.67	6.3E-01	AE002329.2	NT	Chlamydia muridarum, section 59 of 85 of the complete genome
10326	22973	36193	1.47	6.3E-01	Z73003.1	NT	S.cerevisiae chromosome VII reading frame ORF YGR218w
10427	23073	36294	1.19	6.3E-01	AE000313.1	NT	Escherichia coli K-12 MG1655 section 203 of 400 of the complete genome
10456	23102		0.45	6.3E-01	AW795395.1	EST_HUMAN	PM0-UM0018-130500-003-g12 UM0018 Homo sapiens cDNA
10993	23667	36924	2.21	6.3E-01	AA877715.1	EST_HUMAN	nr09h06.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1161371 3' similar to TR:O02916 O02916 HLARK. ;
11308	23967	37268	9.25	6.3E-01	AI904160.1	EST_HUMAN	CM-BT043-090299-046 BT043 Homo sapiens cDNA
11402	24051	37355	1.66	6.3E-01	P47003	SWISSPROT	HYPOTHETICAL 13.7 KD PROTEIN IN INO1-IDS2 INTERGENIC REGION
11581	24180	37495	1.84	6.3E-01	P36073	SWISSPROT	HYPOTHETICAL 15.3 KD PROTEIN IN VMA12-APN1 INTERGENIC REGION
11988	25355	30607	4.37	6.3E-01	9910293	NT	Mus musculus keratin complex 2, gene 6g (Krt2-6g), mRNA
12078	24587		1.45	6.3E-01	AF105227.1	NT	Homo sapiens 3'-phosphoadenosine 5'-phosphosulfate synthetase (PAPSS) mRNA, complete cds
12283	25272		2.93	6.3E-01	X83528.1	NT	C.limicola pecD gene
5780	18571	31499	2.31	6.2E-01	Q10135	SWISSPROT	HYPOTHETICAL 142.5 KD PROTEIN C23E2.02 IN CHROMOSOME I
7394	20073		3.44	6.2E-01	AF022253.1	NT	Mus musculus calcium-sensing receptor related protein 4 (Casr-rs4) mRNA, partial cds
7443	25114	33209	1.33	6.2E-01	AL021127.2	NT	Mus musculus chromosome X contigA; putative Magea9 gene, Caltractin, NAD(P) steroid dehydrogenase and Zinc finger protein 185
8200	20894	34031	4.52	6.2E-01	H72255.1	EST_HUMAN	ys01e08.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:213542 3'
8755	21447	34595	0.52	6.2E-01	AF034411.1	NT	Lycopersicon esculentum cytosolic Cu,Zn superoxide dismutase (Sod) gene, partial cds; and dehydroquinate dehydratase/shikimate:NADP oxidoreductase gene, complete cds
9349	20420	33540	1.55	6.2E-01	BE562687.1	EST_HUMAN	601336146F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3690010 5'
9410	22072		2.55	6.2E-01	M24461.1	NT	Human pulmonary surfactant-associated protein SP-B (SFTP3) mRNA, complete cds
9978	22626	35834	6.2	6.2E-01	AL161511.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 23
10121	22769	35982	0.5	6.2E-01	11420793	NT	Homo sapiens potassium voltage-gated channel, Shab-related subfamily, member 1 (KCNB1), mRNA
10121	22769	35983	0.5	6.2E-01	11420793	NT	Homo sapiens potassium voltage-gated channel, Shab-related subfamily, member 1 (KCNB1), mRNA

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10434	23080	36305	5.2	6.2E-01	P27410	SWISSPROT	NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE ; THIOL PROTEASE P3C ; HELICASE (2C LIKE PROTEIN); COAT PROTEIN]
10434	23080	36306	5.2	6.2E-01	P27410	SWISSPROT	NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE ; THIOL PROTEASE P3C ; HELICASE (2C LIKE PROTEIN); COAT PROTEIN]
2393	15114		4.38	6.1E-01	6678076	NT	Mus musculus secreted acidic cysteine rich glycoprotein (Sparc), mRNA
5449	18248	31137	1.15	6.1E-01	M59940.1	NT	Caenorhabditis elegans N2 CeMyoD (hlh-1) alternatively spliced genes, complete cds
6770	19514	32540	4.02	6.1E-01	M64733.1	NT	Rat TRPM-2 gene, complete cds
6770	19514	32541	4.02	6.1E-01	M64733.1	NT	Rat TRPM-2 gene, complete cds
6920	19656	32702	0.64	6.1E-01	AW105653.1	EST_HUMAN	xd50h03.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2597237 3' similar to gb:X12671_ma1 HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN A1 (HUMAN);
7005	19697	32751	0.72	6.1E-01	Q63769	SWISSPROT	SUSHI REPEAT-CONTAINING PROTEIN SRPX PRECURSOR (DRS PROTEIN) (DOWN-REGULATED BY V-SRC)
8132	20826	33962	3.27	6.1E-01	AF033535.1	NT	Arabidopsis thaliana putative zinc transporter (ZIP1) mRNA, complete cds
8694	21386	34528	1.09	6.1E-01	11431065	NT	Homo sapiens mitogen-activated protein kinase kinase kinase kinase 4 (MAP4K4), mRNA
8694	21386	34529	1.09	6.1E-01	11431065	NT	Homo sapiens mitogen-activated protein kinase kinase kinase kinase 4 (MAP4K4), mRNA
9315	21982	35153	18.74	6.1E-01	AF236117.1	NT	Homo sapiens G-protein coupled receptor EDG-7 mRNA, complete cds
9315	21982	35154	18.74	6.1E-01	AF236117.1	NT	Homo sapiens G-protein coupled receptor EDG-7 mRNA, complete cds
9742	22393	35597	0.93	6.1E-01	AE004452.1	NT	Pseudomonas aeruginosa PA01, section 13 of 529 of the complete genome
9946	22594	35797	1.06	6.1E-01	AF119117.1	NT	Homo sapiens dopamine transporter (SLC6A3) gene, complete cds
11738	24331	37655	2.57	6.1E-01	S83182.1	NT	hyaluronan-binding protein=hepatocyte growth factor activator homolog [human, plasma, mRNA, 2408 nt]
11738	24331	37656	2.57	6.1E-01	S83182.1	NT	hyaluronan-binding protein=hepatocyte growth factor activator homolog [human, plasma, mRNA, 2408 nt]
12074	25159	30899	2.28	6.1E-01	AB041350.1	NT	Mus musculus Col4a5 mRNA for type IV collagen alpha 5 chain, complete cds
12694	24977		1.57	6.1E-01	X95287.1	NT	M.mazei orfA, orfB, and orfC of archaeal ABC-transporter system
482	13267	25903	1.24	6.0E-01	D87675.1	NT	Homo sapiens DNA for amyloid precursor protein, complete cds
548	13331		3.09	6.0E-01	5802999	NT	Homo sapiens adaptor-related protein complex 3, mu 2 subunit (CLA20), mRNA
1341	14089	26765	1.91	6.0E-01	AF065253.1	NT	Human respiratory syncytial virus strain CH93-53b attachment protein (G) gene, complete cds
3795	16547	29180	0.92	6.0E-01	AJ233396.1	NT	Viral hemorrhagic septicemia virus N, P, M, G, Nv, L genes, French strain 07-71
4165	16905		1.09	6.0E-01	AF058895.1	NT	Homo sapiens Notch3 (NOTCH3) gene, exons 26, 27, and 28
5199	18007	30628	2	6.0E-01	P20288	SWISSPROT	D(2) DOPAMINE RECEPTOR
5353	18156	30839	2.86	6.0E-01	AW139713.1	EST_HUMAN	UI-H-BI1-aeb-a-10-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone IMAGE:2718619 3'
6445	19213	32210	2.78	6.0E-01	U38813.1	NT	Musca domestica insecticide-susceptible strain voltage-sensitive sodium channel mRNA, complete cds

Page 45 of 536

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6563	19328	32335	0.68	6.0E-01	Q04912	SWISSPROT	MACROPHAGE-STIMULATING PROTEIN RECEPTOR PRECURSOR (MSP RECEPTOR) (P185-RON) (CDW136) (CD136 ANTIGEN)
7254	19938	33013	6.99	6.0E-01	AJ277661.1	NT	Homo sapiens partial LMO1 gene for LIM domain only 1 protein, exon 1
8023	20718	33850	4.39	6.0E-01	P02835	SWISSPROT	SEGMENTATION PROTEIN FUSHI TARAZU
8023	20718	33851	4.39	6.0E-01	P02835	SWISSPROT	SEGMENTATION PROTEIN FUSHI TARAZU
9723	22374	35574	1.61	6.0E-01	AB008193.1	NT	Homo sapiens genes for leukotriene B4 receptor BLT2, leukotriene B4 receptor BLT1, complete cds
10173	22821		1.46	6.0E-01	Q01497	SWISSPROT	PEROXISOMAL MEMBRANE PROTEIN PER9 (PEROXIN-3)
10990	23664	36921	1.49	6.0E-01	AJ131892.1	NT	Gallus gallus mRNA for Hyperion protein, 419 kD isoform
10990	23664	36922	1.49	6.0E-01	AJ131892.1	NT	Gallus gallus mRNA for Hyperion protein, 419 kD isoform
11640	24140	37449	3.77	6.0E-01	AI420623.1	EST_HUMAN	tf08f07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2095621 3'
12354	24758	31060	2.25	6.0E-01	11421663	NT	Homo sapiens nuclear factor (erythroid-derived 2)-like 3 (NFE2L3), mRNA
12455	24824		2.6	6.0E-01	AA706087.1	EST_HUMAN	zj96g05.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:462776 3'
12639	25208	30815	3.04	6.0E-01	9055303	NT	Mus musculus cGMP-inhibited phosphodiesterase (Pde3a), mRNA
12664	25142		2.06	6.0E-01	BE157617.1	EST_HUMAN	RC1-HT0375-030500-015-c03 HT0375 Homo sapiens cDNA
980	13745	26407	1.36	5.9E-01	U32701.1	NT	Haemophilus influenzae Rd section 16 of 163 of the complete genome
3264	16026	28675	2.29	5.9E-01	AL163267.2	NT	Homo sapiens chromosome 21 segment HS21C067
3264	16026	28676	2.29	5.9E-01	AL163267.2	NT	Homo sapiens chromosome 21 segment HS21C067
4196	16937		4.21	5.9E-01	AF162756.1	NT	Rattus norvegicus cenexin 2 mRNA, partial cds
6373	19142	32139	1.55	5.9E-01	AF085440.2	NT	Homo sapiens low density lipoprotein receptor-related protein II (LRP2) gene, exon 1 and partial cds
7166	19853	32922	1.32	5.9E-01	AB023486.1	NT	Homo sapiens gene for histamine H2 receptor, promoter region and complete cds
7296	19979		0.61	5.9E-01	X68801.1	NT	G.gallus gene for skeletal alpha-actinin, exon EF2
7898	20593	33725	0.46	5.9E-01	D90911.1	NT	Synechocystis sp. PCC6803 complete genome, 13/27, 1576593-1719643
8536	21228	34370	0.48	5.9E-01	D12922.1	NT	Legionella pneumophila gene for iron superoxide dismutase, complete cds
9443	22120	35299	0.82	5.9E-01	AF063204.2	NT	Chlamydia trachomatis strain K/UW31/Cx major outer membrane protein (omp1) gene, complete cds
9813	22464		0.74	5.9E-01	P06463	SWISSPROT	E6 PROTEIN
10088	22738	35951	1.15	5.9E-01	P55284	SWISSPROT	VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5)
10569	23264	36502	2.5	5.9E-01	Q9X0I3	SWISSPROT	THYMIDYLATE KINASE (DTMP KINASE)
10576	23271	36507	1.72	5.9E-01	AF197944.1	NT	Xenopus laevis receptor protein tyrosine phosphatase delta (XPTP-D) mRNA, complete cds
10881	23561	36808	2.91	5.9E-01	AW937175.1	EST_HUMAN	PM1-DT0041-190100-002-h03 DT0041 Homo sapiens cDNA
11149	23816	37099	1.95	5.9E-01	AF064626.1	NT	Mus spretus strain SPRET/EI CD48 antigen (Cd48) gene, partial cds
11458	24062	37368	1.56	5.9E-01	P47135	SWISSPROT	JSN1 PROTEIN

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11458	24062	37369	1.56	5.9E-01	P47135	SWISSPROT	JSN1 PROTEIN
12021	24549	31109	2	5.9E-01	L42320.1	NT	Oryctolagus cuniculus alpha 1 anti-trypsin (alpha 1 AT) gene, promoter region
12252	24698		4.35	5.9E-01	AB017705.1	NT	Aspergillus oryzae pyrG gene for orotidine-5'-phosphate decarboxylase, complete cds
12465	24832		5.72	5.9E-01	P34926	SWISSPROT	MICROTUBULE-ASSOCIATED PROTEIN 1A [CONTAINS: MAP1 LIGHT CHAIN LC2]
1902	14639	27348	1.36	5.8E-01	P40472	SWISSPROT	SIM1 PROTEIN
2569	15283	28021	1.01	5.8E-01	7305230	NT	Mus musculus low-density lipoprotein B (Ldlb), mRNA
4478	17213	29838	4.37	5.8E-01	AB009077.1	NT	Vigna radiata mRNA for proton pyrophosphatase, complete cds
5290	18095		0.82	5.8E-01	AE002152.1	NT	Ureaplasma urealyticum section 53 of 59 of the complete genome
5444	18243	31131	0.62	5.8E-01	Q10699	SWISSPROT	POTENTIAL 5'-3' EXONUCLEASE
6091	18869	31835	1.09	5.8E-01	D78659.1	EST_HUMAN	HUM500E06B Human placenta polyA+ (TFujiwara) Homo sapiens cDNA clone GEN-500E06 5'
6220	18994	31970	0.66	5.8E-01	D50601.1	NT	Shigella sonnei DNA for 26 ORFs, complete cds
6715	19630		2.48	5.8E-01	S65091.1	NT	cyclic AMP-regulated phosphoprotein [rats, mRNA, 1030 nt]
7787	20482		2.61	5.8E-01	H41571.1	EST_HUMAN	yn91b03.s1 Soares adult brain N2b5HB55Y Homo sapiens cDNA clone IMAGE:175757 3' similar to gb:S78187 M-PHASE INDUCER PHOSPHATASE 2 (HUMAN);
7985	20680	33805	0.64	5.8E-01	AI280051.1	EST_HUMAN	qh85d10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1853779 3'
7985	20680	33806	0.64	5.8E-01	AI280051.1	EST_HUMAN	qh85d10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1853779 3'
8090	20784	33914	3.41	5.8E-01	P14328	SWISSPROT	SPORE COAT PROTEIN SP96
8090	20784	33915	3.41	5.8E-01	P14328	SWISSPROT	SPORE COAT PROTEIN SP96
8789	21481	34628	8.97	5.8E-01	AJ270774.1	NT	Homo sapiens partial TCF-4 gene for T-cell transcription factor-4, exons 6-11
8871	21562	34707	0.99	5.8E-01	Q27368	SWISSPROT	TRANSCRIPTION FACTOR E2F
8872	21563	34708	0.51	5.8E-01	Q20471	SWISSPROT	PUTATIVE CASEIN KINASE I F46F2.2 IN CHROMOSOME X
9496	22149		0.81	5.8E-01	BF031606.1	EST_HUMAN	601557774F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3827298 5'
10911	23591	36837	7.56	5.8E-01	AJ243213.1	NT	Homo sapiens partial 5-HT4 receptor gene, exons 2 to 5
10962	23638		3.97	5.8E-01	BF700092.1	EST_HUMAN	602127577F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4284403 5'
11089	23759		1.99	5.8E-01	BF700092.1	EST_HUMAN	602127577F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4284403 5'
1480	14227	26912	1.12	5.7E-01	P06727	SWISSPROT	APOLIPOPROTEIN A-IV PRECURSOR (APO-AIV)
1480	14227	26913	1.12	5.7E-01	P06727	SWISSPROT	APOLIPOPROTEIN A-IV PRECURSOR (APO-AIV)
3038	15804		0.69	5.7E-01	6755253	NT	Mus musculus plasmacytoma variant translocation 1 (Pvt1), mRNA
3217	15980	28631	1.62	5.7E-01	Q9WTJ2	SWISSPROT	PUTATIVE TRANSCRIPTION FACTOR OVO-LIKE 1 (MOVO1) (MOVO1A)
3495	16251		2.82	5.7E-01	AB033503.1	NT	Populus euramericana peacs-2 mRNA for 1-aminocyclopropane-1-carboxylate synthase, complete cds
6262	19036	32011	5.13	5.7E-01	BF035413.1	EST_HUMAN	601454962F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3858590 5'
6611	19374	32388	0.81	5.7E-01	AA194201.1	EST_HUMAN	zr38c06.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:665674 5'
6763	17932	30568	1.33	5.7E-01	AL111440.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
7664	20328	33438	2.14	5.7E-01	P00373	SWISSPROT	PYRROLINE-5-CARBOXYLATE REDUCTASE (P5CR) (P5C REDUCTASE)
7870	20565		0.5	5.7E-01	AJ251835.1	NT	Mus musculus Kcnq1, Litpc5, Mash2, Tapa-1, Tssc4 and Tssc6 genes, alternative transcripts
8279	20973		0.47	5.7E-01	AI065061.1	EST_HUMAN	HA0895 Human fetal liver cDNA library Homo sapiens cDNA
9699	22350	35544	1.19	5.7E-01	AL161532.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 32
9699	22350	35545	1.19	5.7E-01	AL161532.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 32
10475	23121	36351	0.72	5.7E-01	BF540962.1	EST_HUMAN	602067712F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:4066610 5'
11983	24524		1.52	5.7E-01	BE715051.1	EST_HUMAN	MR3-HT0736-180700-003-a02 HT0736 Homo sapiens cDNA
12658	24958		3.01	5.7E-01	BE959722.2	EST_HUMAN	601654814R1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:3839763 3'
3357	16117	28772	1.3	5.6E-01	AB018283.2	NT	Homo sapiens mRNA for KIAA0740 protein, partial cds
3357	16117	28773	1.3	5.6E-01	AB018283.2	NT	Homo sapiens mRNA for KIAA0740 protein, partial cds
3863	16613	29252	0.97	5.6E-01	AL161501.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 13
4215	16956	29578	0.74	5.6E-01	D83135.1	NT	Chicken TBP gene, exon8, complete cds
8702	21394	34541	4.01	5.6E-01	AV684703.1	EST_HUMAN	AV684703 GKC Homo sapiens cDNA clone GKCFSF05 5'
8702	21394	34542	4.01	5.6E-01	AV684703.1	EST_HUMAN	AV684703 GKC Homo sapiens cDNA clone GKCFSF05 5'
9275	22029	35199	1.08	5.6E-01	AB038782.1	NT	Homo sapiens MUC3A gene for intestinal mucin, partial cds
11884	24457		2.57	5.6E-01	BE888280.1	EST_HUMAN	601514007F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3915457 5'
11897	24535	37272	1.63	5.6E-01	AA493535.1	EST_HUMAN	ng75g10.s1 NCI_CGAP_Pr6 Homo sapiens cDNA clone IMAGE:940674 similar to contains element PTR7 repetitive element ;
12352	16613	29252	1.69	5.6E-01	AL161501.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 13
12379	24776		2.7	5.6E-01	P50505	SWISSPROT	HIGH AFFINITY POTASSIUM TRANSPORTER
12773	25027		4.26	5.6E-01	BF573829.1	EST_HUMAN	602132029F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4271334 5'
1189	13941	26606	0.85	5.5E-01	8393912	NT	Rattus norvegicus Propionyl Coenzyme A carboxylase, beta polypeptide (Pccb), mRNA
2705	15412	28149	6.93	5.5E-01	P03341	SWISSPROT	GAG POLYPROTEIN [CONTAINS: INNER COAT PROTEIN P12; CORE PROTEIN P15; CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10]
2705	15412	28150	6.93	5.5E-01	P03341	SWISSPROT	GAG POLYPROTEIN [CONTAINS: INNER COAT PROTEIN P12; CORE PROTEIN P15; CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10]
2919	15685	28330	1	5.5E-01	5902085	NT	Homo sapiens superkiller viralicidic activity 2 (S. cerevisiae homolog)-like (SKIV2L), mRNA
3062	15828		1.55	5.5E-01	H46219.1	EST_HUMAN	yo18a10.s1 Soares adult brain N2b5HB55Y Homo sapiens cDNA clone IMAGE:178266 3'
3228	15991	28644	4.22	5.5E-01	AF227240.1	NT	Rabbit oral papillomavirus, complete genome
3678	16431	29073	1.7	5.5E-01	P48755	SWISSPROT	FOS-RELATED ANTIGEN-1
5082	17801	30419	1.79	5.5E-01	U69097.1	NT	Bos taurus MHC class II beta-chain BoLA-DIB1 gene, partial cds
7187	19873		0.65	5.5E-01	AB015596.1	NT	Carassius auratus gene for gonadotropin II beta subunit, complete cds
8348	21041	34178	1.04	5.5E-01	AI791768.1	EST_HUMAN	or82c01.y5 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1602336 5'
9667	22319		0.7	5.5E-01	U88415.1	NT	Crimean-Congo hemorrhagic fever virus strain SPU 415/85 nucleoprotein gene, complete cds

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10276	22924	36138	0.96	5.5E-01	T05047.1	EST_HUMAN	EST02935 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBCQ35
11087	23757	37033	1.65	5.5E-01	BF129507.1	EST_HUMAN	601811077R1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4054003 3'
140	12955	25597	4.91	5.4E-01	7657266	NT	Homo sapiens KIAA0929 protein Msx2 interacting nuclear target (MINT) homolog (KIAA0929), mRNA
140	12955	25598	4.91	5.4E-01	7657266	NT	Homo sapiens KIAA0929 protein Msx2 interacting nuclear target (MINT) homolog (KIAA0929), mRNA
571	13352	25980	1.16	5.4E-01	AF232006.1	NT	Pseudomonas syringae pv. tomato strain DC3000 AvrE (avrE), HrpW (hrpW), and GstA (gstA) genes, complete cds; and unknown genes
571	13352	25981	1.16	5.4E-01	AF232006.1	NT	Pseudomonas syringae pv. tomato strain DC3000 AvrE (avrE), HrpW (hrpW), and GstA (gstA) genes, complete cds; and unknown genes
1248	13997	26664	3.41	5.4E-01	AW896087.1	EST_HUMAN	QV4-NN0040-070400-160-c04 NN0040 Homo sapiens cDNA
2099	14830		3.43	5.4E-01	AE002247.2	NT	Chlamydophila pneumoniae AR39, section 74 of 94 of the complete genome
2252	14980	27719	1.91	5.4E-01	AJ276682.1	NT	Drosophila melanogaster mRNA for 15,15' beta carotene dioxygenase (beta-diox gene)
5066	17785	30402	0.92	5.4E-01	M74439.1	NT	Rattus rattus UDP glucuronosyltransferase gene, complete cds
5571	18368	31278	0.74	5.4E-01	AW842327.1	EST_HUMAN	PM2-CN0030-030200-003-c10 CN0030 Homo sapiens cDNA
6098	18876	31845	0.83	5.4E-01	AB025017.1	NT	Rattus norvegicus gene for TIS11, complete cds
6928	19684	32710	0.87	5.4E-01	BE966592.2	EST_HUMAN	601660276R1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906090 3'
7235	19920	32993	0.81	5.4E-01	Z21619.1	NT	S.cerevisiae RIB3 gene encoding DBP synthase
7235	19920	32994	0.81	5.4E-01	Z21619.1	NT	S.cerevisiae RIB3 gene encoding DBP synthase
7237	19922	32997	1.48	5.4E-01	Q64428	SWISSPROT	MITOCHONDRIAL TRIFUNCTIONAL ENZYME ALPHA SUBUNIT PRECURSOR (TP-ALPHA) [INCLUDES: LONG-CHAIN ENOYL-COA HYDRATASE ; LONG CHAIN 3-HYDROXYACYL-COA DEHYDROGENASE]
9890	22540		2.09	5.4E-01	BF672536.1	EST_HUMAN	602076545F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4243690 5'
11015	23687	36948	2.87	5.4E-01	P36858	SWISSPROT	NITRATE REDUCTASE [NADPH] (NR)
11621	24218	37541	3.08	5.4E-01	Q60675	SWISSPROT	LAMININ ALPHA-2 CHAIN PRECURSOR (LAMININ M CHAIN) (MEROSIN HEAVY CHAIN)
11621	24218	37542	3.08	5.4E-01	Q60675	SWISSPROT	LAMININ ALPHA-2 CHAIN PRECURSOR (LAMININ M CHAIN) (MEROSIN HEAVY CHAIN)
11944	24499		3.5	5.4E-01	AI858398.1	EST_HUMAN	wl37g04.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2427126 3' similar to gb:M13452 LAMIN A (HUMAN);
503	13287	25921	1.54	5.3E-01	AF019413.1	NT	Homo sapiens HLA class III region containing tenascin X (tenascin-X) gene, partial cds; cytochrome P450 21-hydroxylase (CYP21B), complement component C4 (C4B) G11, helicase (SKI2W), RD, complement factor B (Bf), and complement component C2 (C2) genes,>
2136	14866	27596	1.01	5.3E-01	AF113919.1	NT	Brassica oleracea var. capitata phospholipase D2 (PLD2) gene, complete cds
2136	14866	27597	1.01	5.3E-01	AF113919.1	NT	Brassica oleracea var. capitata phospholipase D2 (PLD2) gene, complete cds
2786	15491	28230	6.83	5.3E-01	4506328	NT	Homo sapiens protein tyrosine phosphatase, receptor-type, zeta polypeptide 1 (PTPRZ1) mRNA

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2786	15491	28231	6.83	5.3E-01	4506328	NT	Homo sapiens protein tyrosine phosphatase, receptor-type, zeta polypeptide 1 (PTPRZ1) mRNA
3237	15999	28649	2.74	5.3E-01	AF087658.1	NT	Homo sapiens secreted C-type lectin precursor (LSLCL) gene, complete cds
4187	16928		1.58	5.3E-01	U39687.1	NT	Mycoplasma genitalium section 9 of 51 of the complete genome
5371	18172	30860	1.96	5.3E-01	AI820921.1	EST_HUMAN	zu42h12.y5 Soares ovary tumor NbHOT Homo sapiens cDNA clone IMAGE:740711 5'
5371	18172	30861	1.96	5.3E-01	AI820921.1	EST_HUMAN	zu42h12.y5 Soares ovary tumor NbHOT Homo sapiens cDNA clone IMAGE:740711 5'
5466	18265	31156	0.84	5.3E-01	AA193672.1	EST_HUMAN	zr42g09.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:666112 5'
5466	18265	31157	0.84	5.3E-01	AA193672.1	EST_HUMAN	zr42g09.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:666112 5'
5559	18356	31266	1.82	5.3E-01	BE645620.1	EST_HUMAN	7e73c12.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3288118 3' similar to gb:J02783 PROTEIN DISULFIDE ISOMERASE PRECURSOR (HUMAN);
5559	18356	31267	1.82	5.3E-01	BE645620.1	EST_HUMAN	7e73c12.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3288118 3' similar to gb:J02783 PROTEIN DISULFIDE ISOMERASE PRECURSOR (HUMAN);
8802	21494		1.8	5.3E-01	L01950.2	NT	Roridula gorgonias ribulose 1,5-bisphosphate carboxylase (rbcL) gene, partial cds; chloroplast gene for chloroplast product
8854	21545	34692	0.81	5.3E-01	BF433956.1	EST_HUMAN	7q71c12.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE: 3' similar to contains element MER29 repetitive element ;
8854	21545	34693	0.81	5.3E-01	BF433956.1	EST_HUMAN	7q71c12.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE: 3' similar to contains element MER29 repetitive element ;
10111	22759	35971	0.62	5.3E-01	AI954210.1	EST_HUMAN	wx94b02.x1 NCI_CGAP_Mel15 Homo sapiens cDNA clone IMAGE:2551275 3' similar to SW:COXA_HUMAN P20674 CYTOCHROME C OXIDASE POLYPEPTIDE VA PRECURSOR ;
11550	24149	37460	7.3	5.3E-01	BE566291.1	EST_HUMAN	601339867F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3682168 5'
11789	24379	37709	1.72	5.3E-01	Q05793	SWISSPROT	BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC)
11877	25206		4.03	5.3E-01	AA916053.1	EST_HUMAN	og30e05.s1 NCI_CGAP_Br7 Homo sapiens cDNA clone IMAGE:1441376 3' similar to gb:J02611 APOLIPOPROTEIN D PRECURSOR (HUMAN);
797	13569	26229	18.35	5.2E-01	L20770.1	NT	Drosophila melanogaster helix-loop-helix mRNA, complete cds
1141	13896	26557	8.29	5.2E-01	Q9WV30	SWISSPROT	NUCLEAR FACTOR OF ACTIVATED T CELLS 5 (T CELL TRANSCRIPTION FACTOR NFAT5) (NF-AT5) (REL DOMAIN-CONTAINING TRANSCRIPTION FACTOR NFAT5)
1169	13923	26585	1.77	5.2E-01	AF224492.1	NT	Homo sapiens phospholipid scramblase 1 gene, complete cds
1879	14616		2.35	5.2E-01	AL163285.2	NT	Homo sapiens chromosome 21 segment HS21C085
2142	14872	27605	2.55	5.2E-01	AB018283.2	NT	Homo sapiens mRNA for KIAA0740 protein, partial cds
3117	15882	28521	1.23	5.2E-01	U65942.1	NT	Chlamydophila abortus strain S26/3 POMP91A and POMP90A precursor, genes, complete cds
3231	15994		1	5.2E-01	D73443.1	NT	Azotobacter vinelandii icd gene for isocitrate dehydrogenase, complete cds
3400	16158		1.58	5.2E-01	AL116780.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
3437	16193	28843	2.27	5.2E-01	AA984165.1	EST_HUMAN	am77g05.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone IMAGE:1616504 3'

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
3623	16376		0.76	5.2E-01	AF020269.1	NT	Medicago sativa chloroplast malate dehydrogenase precursor (p1mdh) mRNA, nuclear gene encoding chloroplast protein, complete cds
4568	17303	29930	0.82	5.2E-01	6752947	NT	Mus musculus acetylcholine receptor beta (Acrb), mRNA
4953	17679		1.02	5.2E-01	7106444	NT	Mus musculus vanilloid receptor-like protein 1 (Vrl1), mRNA
5567	18364	31272	0.87	5.2E-01	AA284261.1	EST_HUMAN	zc44d09.T7 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:325169 3'
9630	25126	35474	0.75	5.2E-01	X02218.1	NT	Chicken duplicated genes for histone H2A, H4 and a histone H3 gene
9630	25126	35475	0.75	5.2E-01	X02218.1	NT	Chicken duplicated genes for histone H2A, H4 and a histone H3 gene
9832	22483	35685	0.48	5.2E-01	AA194518.1	EST_HUMAN	zq05b09.r1 Stratagene muscle 937209 Homo sapiens cDNA clone IMAGE:628793 5'
9926	22574	35772	1.35	5.2E-01	AF143952.2	NT	Homo sapiens PELOTA (PELOTA) gene, complete cds
12744	25010		7	5.2E-01	P18516	SWISSPROT	RETINOIC ACID RECEPTOR GAMMA (RAR-GAMMA) (RETINOIC ACID RECEPTOR DELTA) (RAR-DELTA)
603	13381	26013	1.84	5.1E-01	M58509.1	NT	Human adrenodoxin reductase gene, exons 3 to 12
633	13412	26047	4.49	5.1E-01	AJ233944.1	NT	Polyangium vitellinum (strain Pl vt1) 16S rRNA gene
633	13412	26048	4.49	5.1E-01	AJ233944.1	NT	Polyangium vitellinum (strain Pl vt1) 16S rRNA gene
1648	14394		1.09	5.1E-01	X87885.1	NT	R.norvegicus mRNA for mammalian fusca protein
2017	14752		1.29	5.1E-01	BF683095.1	EST_HUMAN	602139319F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4298117 5'
4057	16802	29433	3.86	5.1E-01	AI858495.1	EST_HUMAN	wl39b12.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2427263 3'
4164	16904	29533	2.81	5.1E-01	P96380	SWISSPROT	TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF)
5103	17821	30438	1.01	5.1E-01	U72653.1	NT	Human alpha 1a adrenergic receptor (alpha1a) gene, 5' flanking region
6128	18906	31874	0.67	5.1E-01	BE541068.1	EST_HUMAN	601063606F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3450000 5'
6183	18960		0.93	5.1E-01	AV712326.1	EST_HUMAN	AV712326 DCA Homo sapiens cDNA clone DCAAUF07 5'
6818	19479	32502	1.69	5.1E-01	R80873.1	EST_HUMAN	yi94a09.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:146872 3'
8470	21162	34304	0.63	5.1E-01	AW806881.1	EST_HUMAN	QV4-ST0023-160400-172-e01 ST0023 Homo sapiens cDNA
8470	21162	34305	0.63	5.1E-01	AW806881.1	EST_HUMAN	QV4-ST0023-160400-172-e01 ST0023 Homo sapiens cDNA
9583	22236	35420	4.33	5.1E-01	J05412.1	NT	Human regenerating protein (reg) gene, complete cds
9587	22240	35424	3.14	5.1E-01	W22302.1	EST_HUMAN	65B1 Human retina cDNA Tsp509I-cleaved sublibrary Homo sapiens cDNA not directional
10060	22708	35926	0.89	5.1E-01	M94579.1	NT	Human carboxyl ester lipase (CEL) gene, complete cds
12086	25137		4.26	5.1E-01	BF030207.1	EST_HUMAN	601556863F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3826767 5'
12326	24745		3.55	5.1E-01	BF439982.1	EST_HUMAN	nac51f10.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3406218 3' similar to contains element TAR1 repetitive element ;
2130	14861	27590	1.24	5.0E-01	4885552	NT	Homo sapiens postmeiotic segregation increased 2-like 9 (PMS2L9), mRNA
2130	14861	27591	1.24	5.0E-01	4885552	NT	Homo sapiens postmeiotic segregation increased 2-like 9 (PMS2L9), mRNA

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2140	14870	27601	3.19	5.0E-01	AF008210.1	NT	Buchnera aphidicola genomic fragment containing (chaperone Hsp60) groEL, DNA biosynthesis initiating protein (dnaA), ATP operon (atpCDGAHFEB), and putative chromosome replication protein (gidA) genes, complete cds; and termination factor Rho (rho) gene>
2140	14870	27602	3.19	5.0E-01	AF008210.1	NT	Buchnera aphidicola genomic fragment containing (chaperone Hsp60) groEL, DNA biosynthesis initiating protein (dnaA), ATP operon (atpCDGAHFEB), and putative chromosome replication protein (gidA) genes, complete cds; and termination factor Rho (rho) gene>
3811	16563	29196	1.13	5.0E-01	L38483.1	NT	Rattus norvegicus jagged protein mRNA, complete cds
3854	16604	29241	2.75	5.0E-01	AB033010.1	NT	Homo sapiens mRNA for KIAA1184 protein, partial cds
6547	19312		0.65	5.0E-01	BF576199.1	EST_HUMAN	602132842F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4271939 5'
7562	20232	33334	0.75	5.0E-01	AL161549.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 49
7562	20232	33335	0.75	5.0E-01	AL161549.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 49
8428	21121		1.82	5.0E-01	M92304.1	NT	Xenopus laevis smooth muscle beta-tropomyosin mRNA, complete cds
8569	21261	34399	0.71	5.0E-01	BF107848.1	EST_HUMAN	601823850R1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4043485 3'
9358	20429	33547	2.74	5.0E-01	BF317212.1	EST_HUMAN	601903871F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4136632 5'
9525	22178	35362	1.36	5.0E-01	P35573	SWISSPROT	GLYCOGEN DEBRANCHING ENZYME (GLYCOGEN DEBRANCHER) [INCLUDES: 4-ALPHA-GLUCANOTRANSFERASE (OLIGO-1,4-1,4-GLUCANTRANSFERASE); AMYLO-1,6-GLUCOSIDASE (DEXTRIN 6-ALPHA-D-GLUCOSIDASE)]
9525	22178	35363	1.36	5.0E-01	P35573	SWISSPROT	GLYCOGEN DEBRANCHING ENZYME (GLYCOGEN DEBRANCHER) [INCLUDES: 4-ALPHA-GLUCANOTRANSFERASE (OLIGO-1,4-1,4-GLUCANTRANSFERASE); AMYLO-1,6-GLUCOSIDASE (DEXTRIN 6-ALPHA-D-GLUCOSIDASE)]
10290	22938		1.12	5.0E-01	BE869218.1	EST_HUMAN	601445024F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3849436 5'
12026	24554		4	5.0E-01	AF029215.1	NT	Mus musculus MRC OX-2 antigen homolog gene, exons 2-5, and complete cds
12715	24989		1.86	5.0E-01	AL163302.2	NT	Homo sapiens chromosome 21 segment HS21C102
12726	24997		4.39	5.0E-01	O13961	SWISSPROT	NUCLEAR ENVELOPE PROTEIN CUT11
772	13544	26205	2.43	4.9E-01	BF571462.1	EST_HUMAN	602076649F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4243860 5'
1656	14402	27090	1.54	4.9E-01	AJ243955.1	NT	Xenopus laevis mRNA for c-Jun protein, 1978 BP
1899	14636	27345	1.15	4.9E-01	U40869.1	NT	Cavia porcellus pulmonary surfactant protein A (SP-a) mRNA, complete cds
5321	18124	30783	0.89	4.9E-01	Q61554	SWISSPROT	FIBRILLIN 1 PRECURSOR
5946	18728	31686	3.05	4.9E-01	AF020931.1	NT	Homo sapiens diacylglycerol kinase 3 (DAGK3) gene, exon 10
5946	18728	31687	3.05	4.9E-01	AF020931.1	NT	Homo sapiens diacylglycerol kinase 3 (DAGK3) gene, exon 10
7352	20033	33111	1.61	4.9E-01	AB040051.1	NT	Oryza sativa subsp. japonica mEF-G mRNA for mitochondrial elongation factor G, complete cds
7605	20271	33378	0.84	4.9E-01	Q10606	SWISSPROT	PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE
7605	20271	33379	0.84	4.9E-01	Q10606	SWISSPROT	PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE
8888	21579		1.45	4.9E-01	BF209791.1	EST_HUMAN	601874984F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4102503 5'

Page 52 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9086	21775	34939	0.99	4.9E-01	AW339905.1	EST_HUMAN	hc90c02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2907266 3' similar to TR:O95714 O95714 HERC2. ;
9196	25431		1.96	4.9E-01	10946863	NT	Mus musculus unc13 homolog (C. elegans) 1 (Unc13h1), mRNA
10216	22864	36076	0.84	4.9E-01	AF053980.1	NT	Mus musculus adenylyl cyclase 1 (Adcy1) cDNA, partial cds
10419	23065	36286	0.77	4.9E-01	X90000.1	NT	H.sapiens DNA for BCL7A gene and BCL7A/IGH locus fussion
11925	24486		1.72	4.9E-01	AF176912.1	NT	Homo sapiens neurotrophin-1/B-cell stimulating factor-3 gene, complete cds
12709	25392		6.73	4.9E-01	AA613562.1	EST_HUMAN	nq22e11.s1 NCI_CGAP_Co10 Homo sapiens cDNA clone IMAGE:1144652 3'
4298	17037		0.77	4.8E-01	4504850	NT	Homo sapiens potassium channel, subfamily K, member 5 (TASK-2) (KCNK5) mRNA, and translated products
5420	18219	30930	10.78	4.8E-01	J02987.1	NT	Saccharomyces cerevisiae) sporulation protein (SPO11) gene required for meiotic recombination, complete cds
6579	19342	32356	0.79	4.8E-01	U92882.1	NT	Mus musculus slow skeletal muscle troponin T (Tnnt1) gene, complete cds
6589	19352		3.76	4.8E-01	AA659878.1	EST_HUMAN	nu85f09.s1 NCI_CGAP_Alv1 Homo sapiens cDNA clone IMAGE:1217513
7216	19901		1.99	4.8E-01	5031650	NT	Homo sapiens reproduction 8 (D8S2298E) mRNA
7565	20235	33339	0.78	4.8E-01	AL163209.2	NT	Homo sapiens chromosome 21 segment HS21C009
7661	20325	33434	4.05	4.8E-01	AL161492.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 4
7661	20325	33435	4.05	4.8E-01	AL161492.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 4
7805	20500	33621	1.2	4.8E-01	AI820744.1	EST_HUMAN	yj77f10.y5 Soares breast 2NbHBst Homo sapiens cDNA clone IMAGE:154795 5' similar to contains element MER6 repetitive element ;
9144	21875		0.92	4.8E-01	BE155148.1	EST_HUMAN	PM1-HT0350-201299-004-b04 HT0350 Homo sapiens cDNA
10629	23322		1.88	4.8E-01	X83502.1	NT	S.cerevisiae ORFs from chromosome X
12217	25165		3.04	4.8E-01	AF227565.1	NT	Trypanosoma cruzi transposon VIP II SIRE repeat region
12795	25216		1.66	4.8E-01	AJ132984.1	NT	Chlamydomonas reinhardtii cop gene, exons 1-8
6422	19190	32186	8.41	4.7E-01	BF217173.1	EST_HUMAN	601883880F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4096387 5'
6941	19423	32438	0.94	4.7E-01	AI204374.1	EST_HUMAN	qf72a09.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1755544 3'
7764	20460	33584	0.63	4.7E-01	T11414.1	EST_HUMAN	hbc811 Human pancreatic islet Homo sapiens cDNA clone hbc811 5'end
7764	20460	33585	0.63	4.7E-01	T11414.1	EST_HUMAN	hbc811 Human pancreatic islet Homo sapiens cDNA clone hbc811 5'end
8974	21664	34816	0.52	4.7E-01	6981501	NT	Rattus norvegicus Spermine binding protein (Sbp), mRNA
10751	23436		6.11	4.7E-01	AF102673.1	NT	Influenza A virus isolate hk51697 hemagglutinin (HA) gene, partial cds
11022	23694	36957	2.2	4.7E-01	U41069.1	NT	Human collagen alpha2(XI) (COL11A2) gene, exons 6 through 16, and partial cds
11252	23914	37206	1.61	4.7E-01	BF529658.1	EST_HUMAN	602043889F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4181303 5'
11349	24039	37342	1.7	4.7E-01	AW889448.1	EST_HUMAN	RC6-NT0029-240400-011-E08 NT0029 Homo sapiens cDNA
12116	24609		1.52	4.7E-01	BE887763.1	EST_HUMAN	601511333F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3912488 5'
12237	24689		1.51	4.7E-01	AW341561.1	EST_HUMAN	hd11c08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2909198 3'

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
12813	25055		1.63	4.7E-01	AP000007.1	NT	Pyrococcus horikoshii OT3 genomic DNA, 1485001-1738505 nt. position (7/7)
12817	25300		1.38	4.7E-01	6679502	NT	Mus musculus proteasome (prosome, macropain) 26S subunit, ATPase 3 (Psmc3), mRNA
3726	16479	29116	1.57	4.6E-01	BF693300.1	EST_HUMAN	602081103F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4245481 5'
3726	16479	29117	1.57	4.6E-01	BF693300.1	EST_HUMAN	602081103F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4245481 5'
5333	18136	30795	1	4.6E-01	BF313593.1	EST_HUMAN	601900234F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4129472 5'
5333	18136	30796	1	4.6E-01	BF313593.1	EST_HUMAN	601900234F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4129472 5'
5385	18185	30875	3.11	4.6E-01	Q90843	SWISSPROT	INTERFERON REGULATORY FACTOR 3 (IRF-3)
5385	18185	30876	3.11	4.6E-01	Q90843	SWISSPROT	INTERFERON REGULATORY FACTOR 3 (IRF-3)
5459	18258	31148	1.84	4.6E-01	BE734781.1	EST_HUMAN	601568755F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3843637 5'
5472	18271	31163	2.17	4.6E-01	AI247679.1	EST_HUMAN	qh59h02.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1849011 3' similar to TR:O15338 O15338 BUTYROPHILIN. ;
5472	18271	31164	2.17	4.6E-01	AI247679.1	EST_HUMAN	qh59h02.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:1849011 3' similar to TR:O15338 O15338 BUTYROPHILIN. ;
5480	18279	31175	1.6	4.6E-01	P20050	SWISSPROT	MEIOSIS SPECIFIC PROTEIN HOP1
5560	18357		0.96	4.6E-01	AF212124.1	NT	Anolis schwartzi cytochrome b gene, partial cds; mitochondrial gene for mitochondrial product
5645	18440		0.77	4.6E-01	BE817247.1	EST_HUMAN	PM0-BN0260-120600-001-F07 BN0260 Homo sapiens cDNA
5809	18598	31526	0.59	4.6E-01	D26215.1	NT	Unidentified soil bacteria 16S rRNA gene encoding 16S ribosomal RNA
6163	18940	31911	1.21	4.6E-01	AE000894.1	NT	Methanobacterium thermoautotrophicum from bases 1165751 to 1176238 (section 100 of 148) of the complete genome
6669	19586	32620	3.2	4.6E-01	U62332.1	NT	Emericella nidulans NEMPA (nempA) gene, mitochondrial gene encoding putative mitochondrial protein, complete cds
6669	19586	32621	3.2	4.6E-01	U62332.1	NT	Emericella nidulans NEMPA (nempA) gene, mitochondrial gene encoding putative mitochondrial protein, complete cds
7131	25105	32884	0.57	4.6E-01	L07320.1	NT	Murine cytomegalovirus e1 protein gene, complete cds
7629	20295	33403	0.91	4.6E-01	AA493577.1	EST_HUMAN	nh04h05.s1 NCI_CGAP_Thy1 Homo sapiens cDNA clone IMAGE:943353 similar to contains Alu repetitive element;contains element L1 repetitive element ;
7658	20322		0.59	4.6E-01	Q90069	SWISSPROT	GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER COMPONENT PROTEINASE (HC-PRO); PROTEIN P3; 6 KD PROTEIN 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KD PROTEIN 2 (6K2); GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NI>
8219	20913	34049	10.11	4.6E-01	BF697399.1	EST_HUMAN	602130953F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4287828 5'
9201	21870	35035	1.11	4.6E-01	P55202	SWISSPROT	ATRIAL NATRIURETIC PEPTIDE RECEPTOR B PRECURSOR (ANP-B) (ANPRB) (GC-B) (GUANYLATE CYCLASE)

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source.	Top Hit Descriptor
9201	21870	35036	1.11	4.6E-01	P55202	SWISSPROT	ATRIAL NATRIURETIC PEPTIDE RECEPTOR B PRECURSOR (ANP-B) (ANPRB) (GC-B) (GUANYLATE CYCLASE)
9876	22526	35720	1.64	4.6E-01	AI915634.1	EST_HUMAN	wg73e12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:2370766 3'
9876	22526	35721	1.64	4.6E-01	AI915634.1	EST_HUMAN	wg73e12.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:2370766 3'
10912	23592		2.3	4.6E-01	P98163	SWISSPROT	PUTATIVE VITELLOGENIN RECEPTOR PRECURSOR (YL)
10922	23602	36850	10.22	4.6E-01	BE185449.1	EST_HUMAN	IL5-HT0730-100500-075-g05 HT0730 Homo sapiens cDNA
10922	23602	36851	10.22	4.6E-01	BE185449.1	EST_HUMAN	IL5-HT0730-100500-075-g05 HT0730 Homo sapiens cDNA
11450	23217	36449	5.32	4.6E-01	AF019369.1	NT	Human thiopurine methyltransferase (TPMT) gene, exon 10 and complete cds
11450	23217	36450	5.32	4.6E-01	AF019369.1	NT	Human thiopurine methyltransferase (TPMT) gene, exon 10 and complete cds
12163	24645		1.77	4.6E-01	D53316.1	EST_HUMAN	HUM105F03B Clontech human fetal brain polyA+ mRNA (#6535) Homo sapiens cDNA clone GEN-105F03 5'
1904	14641	27350	1.43	4.5E-01	AE001931.1	NT	Deinococcus radiodurans R1 section 68 of 229 of the complete chromosome 1
1904	14641	27351	1.43	4.5E-01	AE001931.1	NT	Deinococcus radiodurans R1 section 68 of 229 of the complete chromosome 1
2873	15640	28284	4.5	4.5E-01	AA677086.1	EST_HUMAN	zj55d02.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:454179 3'
3312	16072	28722	4.58	4.5E-01	Q05793	SWISSPROT	BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC)
3372	16131	28787	1.07	4.5E-01	AF126378.1	NT	Mus musculus DNA polymerase epsilon catalytic subunit (Pole) gene, exons 2 through 12
4007	16753		0.95	4.5E-01	Q28247	SWISSPROT	COLLAGEN ALPHA 5(IV) CHAIN
4055	16800	29431	0.88	4.5E-01	AI708908.1	EST_HUMAN	as96e09.x1 Barstead aorta HPLRB6 Homo sapiens cDNA clone IMAGE:2353480 3'
4155	17887		4.25	4.5E-01	AW873495.1	EST_HUMAN	ho60g02.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:3041810 3'
4890	17617	30236	1.1	4.5E-01	BE963445.2	EST_HUMAN	601657225R1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3866023 3'
5461	18260	31151	1.49	4.5E-01	AW608814.1	EST_HUMAN	QV2-PT0012-140100-031-c09 PT0012 Homo sapiens cDNA
6510	19275		1.45	4.5E-01	Q00956	SWISSPROT	COAT PROTEIN
7312	19995	33073	1.27	4.5E-01	M37036.1	NT	Rat nucleolar proteins B23.1 and B23.2
7509	20180	33273	2.54	4.5E-01	AI858849.1	EST_HUMAN	wl32e02.x1 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2426618 3' similar to TR:Q92923 Q92923 SWI/SNF COMPLEX 170 KDA SUBUNIT. ;
7621	20287	33396	0.65	4.5E-01	P50070	SWISSPROT	DNA PRIMASE
8206	20900		0.86	4.5E-01	M32661.1	NT	D.melanogaster Shaw2 protein mRNA, complete cds
8302	20996	34134	3.5	4.5E-01	AI648596.1	EST_HUMAN	tz56g11.x1 NCI_CGAP_Ov35 Homo sapiens cDNA clone IMAGE:2292644 3'
8457	21149	34292	0.83	4.5E-01	Q52728	SWISSPROT	POLY-BETA-HYDROXYBUTYRATE POLYMERASE (POLY(3-HYDROXYBUTYRATE) POLYMERASE) (PHB POLYMERASE) (PHB SYNTHASE) (POLY(3-HYDROXYALKANOATE) POLYMERASE) (PHA POLYMERASE) (PHA SYNTHASE) (POLYHYDROXYALKANOIC ACID SYNTHASE)
8680	21372		2.34	4.5E-01	11444786	NT	Homo sapiens hypothetical protein DKFZp547G183 (DKFZp547G183), mRNA

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8897	21588	34728	0.86	4.5E-01	AE000218.1	NT	Escherichia coli K-12 MG1655 section 108 of 400 of the complete genome
9840	22491		1.02	4.5E-01	9630816	NT	Bombyx mori nuclear polyhedrosis virus, complete genome
10392	23038	36254	24.62	4.5E-01	M86006.1	EST_HUMAN	EST02531 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBCY17
10392	23038	36255	24.62	4.5E-01	M86006.1	EST_HUMAN	EST02531 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBCY17
10772	23455	36699	2.15	4.5E-01	AW591271.1	EST_HUMAN	xo14h01.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2703985 3' similar to SW:INT6_MOUSE Q64252 VIRAL INTEGRATION SITE PROTEIN INT-6. [1] ; '
11217	23880		1.52	4.5E-01	AV719382.1	EST_HUMAN	AV719382 GLC Homo sapiens cDNA clone GLCCED12 5'
11895	25384		3.52	4.5E-01	BE871461.1	EST_HUMAN	601449201F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852961 5'
12540	24880		1.58	4.5E-01	BF337531.1	EST_HUMAN	602035275F1 NCI_CGAP_Bm64 Homo sapiens cDNA clone IMAGE:4183290 5'
12611	24918		3.37	4.5E-01	11422099	NT	Homo sapiens testis-specific kinase 2 (TESK2), mRNA
2388	15109	27847	3.39	4.4E-01	P49765	SWISSPROT	VASCULAR ENDOTHELIAL GROWTH FACTOR B PRECURSOR (VEGF-B) (VEGF RELATED FACTOR)
3310	16070	28719	1.29	4.4E-01	AF058790.1	NT	Rattus norvegicus SynGAP-b mRNA, complete cds
3310	16070	28720	1.29	4.4E-01	AF058790.1	NT	Rattus norvegicus SynGAP-b mRNA, complete cds
3313	16073	28723	2.92	4.4E-01	BF056726.1	EST_HUMAN	7j91d02.y1 NCI_CGAP_Br16 Homo sapiens cDNA clone IMAGE:3393795 5'
4209	16950		1.88	4.4E-01	BE378707.1	EST_HUMAN	601237139F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3609393 5'
5334	18137	30797	1.2	4.4E-01	P04929	SWISSPROT	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR
5334	18137	30798	1.2	4.4E-01	P04929	SWISSPROT	HISTIDINE-RICH GLYCOPROTEIN PRECURSOR
5602	18397	31309	1.59	4.4E-01	S65019.1	NT	mucin [rats, Sprague-Dawley, sulfur-dioxide-treated tracheal epithelium, mRNA Partial, 390 nt]
5619	18415	31328	2	4.4E-01	AV720408.1	EST_HUMAN	AV720408 GLC Homo sapiens cDNA clone GLCCSC12 5'
5864	18651	31591	1.46	4.4E-01	AI198413.1	EST_HUMAN	qi62h11.x1 NCI_CGAP_Bm25 Homo sapiens cDNA clone IMAGE:1861125 3' similar to TR:Q29168 Q29168 UNKNOWN PROTEIN ;
5864	18651	31592	1.46	4.4E-01	AI198413.1	EST_HUMAN	qi62h11.x1 NCI_CGAP_Bm25 Homo sapiens cDNA clone IMAGE:1861125 3' similar to TR:Q29168 Q29168 UNKNOWN PROTEIN ;
6146	18923	31894	1.78	4.4E-01	AW080795.1	EST_HUMAN	xc27e08.x1 NCI_CGAP_Co18 Homo sapiens cDNA clone IMAGE:2585510 3' similar to TR:O95154 O95154 AFLATOXIN B1-ALDEHYDE REDUCTASE. ;
6236	19010		1.42	4.4E-01	AA776132.1	EST_HUMAN	ae85d11.s1 Stratagene schizo brain S11 Homo sapiens cDNA clone IMAGE:970965 3' similar to gb:M16038 TYROSINE-PROTEIN KINASE LYN (HUMAN);
7297	19980	33056	1.04	4.4E-01	AE000571.1	NT	Helicobacter pylori 26695 section 49 of 134 of the complete genome
7723	25119		0.6	4.4E-01	AE001188.1	NT	Treponema pallidum section 4 of 87 of the complete genome
7740	20436		9.71	4.4E-01	Z11679.1	NT	S.tuberosum mRNA for induced stolon tip protein (partial)
8661	21353	34500	0.84	4.4E-01	AA056427.1	EST_HUMAN	zl69a03.s1 Stratagene colon (#937204) Homo sapiens cDNA clone IMAGE:509836 3'
9049	21738	34896	0.7	4.4E-01	AF112540.1	NT	HIV-1 isolate 08107v6 from USA, envelope glycoprotein (env) gene, partial cds

Page 56 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9082	21771	34934	0.57	4.4E-01	AW612578.1	EST_HUMAN	hh05c08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2954222 3' similar to SW:MSH6_HUMAN P52701 DNA MISMATCH REPAIR PROTEIN MSH6 ;
9190	21860	35025	1.24	4.4E-01	O62836	SWISSPROT	ZINC FINGER X-CHROMOSOMAL PROTEIN
9862	22512	35709	2.19	4.4E-01	AI268650.1	EST_HUMAN	qo39f09.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1910921 3'
9863	22513		1.88	4.4E-01	P28922	SWISSPROT	GLYCOPROTEIN B PRECURSOR (GLYCOPROTEIN 14)
9997	22645	35857	4.31	4.4E-01	P35590	SWISSPROT	TYROSINE-PROTEIN KINASE RECEPTOR TIE-1 PRECURSOR
10273	22921	36132	1.33	4.4E-01	S76404.1	NT	beta -HKA=H,K-ATPase beta-subunit [rats, Genomic, 8983 nt, segment 2 of 2]
10273	22921	36133	1.33	4.4E-01	S76404.1	NT	beta -HKA=H,K-ATPase beta-subunit [rats, Genomic, 8983 nt, segment 2 of 2]
12148	24635	31095	3.44	4.4E-01	6677874	NT	Mus musculus sodium channel, type X, alpha polypeptide (Scn10a), mRNA
12579	24903	31000	3.35	4.4E-01	9627742	NT	Autographa californica nucleopolyhedrovirus, complete genome
12683	24971		1.91	4.4E-01	P54725	SWISSPROT	UV EXCISION REPAIR PROTEIN PROTEIN RAD23 HOMOLOG A (HHR23A)
12766	25152		1.43	4.4E-01	AW363338.1	EST_HUMAN	RC2-CT0320-281199-012-c07 CT0320 Homo sapiens cDNA
402	13187	25835	2.17	4.3E-01	AF155218.1	NT	Callithrix jacchus MW/LW opsin gene, upstream flanking region
402	13187	25836	2.17	4.3E-01	AF155218.1	NT	Callithrix jacchus MW/LW opsin gene, upstream flanking region
2875	15642		1.64	4.3E-01	AW935269.1	EST_HUMAN	CM2-DT0003-010200-077-c01 DT0003 Homo sapiens cDNA
3056	15822	28468	0.75	4.3E-01	AW999477.1	EST_HUMAN	MR0-BN0070-270300-008-g04 BN0070 Homo sapiens cDNA
4131	16873	29501	1.29	4.3E-01	J00306.1	NT	Human somatostatin I gene and flanks
4374	13187	25835	1.18	4.3E-01	AF155218.1	NT	Callithrix jacchus MW/LW opsin gene, upstream flanking region
4374	13187	25836	1.18	4.3E-01	AF155218.1	NT	Callithrix jacchus MW/LW opsin gene, upstream flanking region
4902	17629		1.19	4.3E-01	AL161502.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 14
5280	18085	30742	0.8	4.3E-01	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
5280	18085	30743	0.8	4.3E-01	P48634	SWISSPROT	LARGE PROLINE-RICH PROTEIN BAT2 (HLA-B-ASSOCIATED TRANSCRIPT 2)
5798	18589	31516	1.59	4.3E-01	BE181655.1	EST_HUMAN	QV1-HT0638-070500-191-d08 HT0638 Homo sapiens cDNA
5817	18606	31534	2.02	4.3E-01	AF179825.1	NT	Saimiri sciureus olfactory receptor (SSC186) gene, partial cds
6608	19371	32384	4.78	4.3E-01	AJ001678.1	NT	Coturnix coturnix japonica IfnG gene
6689	19606	32646	0.6	4.3E-01	AF075629.1	NT	Equus caballus microsatellite LEX027
6767	19511		0.91	4.3E-01	O33367	SWISSPROT	DNA GYRASE SUBUNIT B
7328	20011		1.88	4.3E-01	BF348001.1	EST_HUMAN	602023134F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4158296 5'
7496	20168	33260	0.61	4.3E-01	U51002.1	NT	Mus musculus Dlx-2 gene, complete cds
8326	21019		2.72	4.3E-01	U97040.1	NT	Methanococcus voltae flagella-related protein C-I (flaC-flaI) genes, complete cds
9154	21885	35053	0.96	4.3E-01	Y14604.1	NT	Erwinia amylovora rcsV gene
9626	22279	35468	2.18	4.3E-01	AW630048.1	EST_HUMAN	hh74e10.y1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2968554 5'
9626	22279	35469	2.18	4.3E-01	AW630048.1	EST_HUMAN	hh74e10.y1 NCI_CGAP_GU1 Homo sapiens cDNA clone IMAGE:2968554 5'

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10128	22776	35990	0.84	4.3E-01	AW170559.1	EST_HUMAN	xn63e05.x1 Soares_NHCeC_cervical_tumor Homo sapiens cDNA clone IMAGE:2698400 3' similar to TR:O00189 O00189 MU-ADAPTIN-RELATED PROTEIN 2. ;
10409	23055	36272	0.5	4.3E-01	H65292.1	EST_HUMAN	yr45b05.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:208209 3'
10849	19606	32646	2.45	4.3E-01	AF075629.1	NT	Equus caballus microsatellite LEX027
11166	23833	37113	1.29	4.3E-01	AW993658.1	EST_HUMAN	RC3-BN0034-290200-013-c12 BN0034 Homo sapiens cDNA
11166	23833	37114	1.29	4.3E-01	AW993658.1	EST_HUMAN	RC3-BN0034-290200-013-c12 BN0034 Homo sapiens cDNA
11745	24336	37662	1.84	4.3E-01	AI874332.1	EST_HUMAN	tz64d04.x1 NCI_CGAP_Ov35 Homo sapiens cDNA clone IMAGE:2293351 3'
12770	25025		2.18	4.3E-01	AJ003022.1	NT	Streptomyces coelicolor whiH gene
1337	15566	26761	1.54	4.2E-01	Q39102	SWISSPROT	CELL DIVISION PROTEIN FTSH HOMOLOG PRECURSOR
1941	14676		1.23	4.2E-01	AA761653.1	EST_HUMAN	nz24a09.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1288696 3'
3596	16349	28990	4.4	4.2E-01	AE003947.1	NT	Xylella fastidiosa, section 93 of 229 of the complete genome
3628	16381	29021	1.41	4.2E-01	AI280338.1	EST_HUMAN	qi94b01.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1879945 3'
3698	17886		0.85	4.2E-01	N81203.1	EST_HUMAN	788iE1 fetal brain cDNA Homo sapiens cDNA clone 788iE1-K similar to R07879, Z40498
3964	16713	29352	0.97	4.2E-01	Q04888	SWISSPROT	SOX-8 PROTEIN
4649	17383	30015	4.88	4.2E-01	AA534093.1	EST_HUMAN	nj69h01.s1 NCI_CGAP_Pr10 Homo sapiens cDNA clone IMAGE:997777 similar to gb:M33600 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR-1 BETA CHAIN (HUMAN);
4731	17463	30100	3.46	4.2E-01	R13467.1	EST_HUMAN	yf77e01.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:28278 5'
5626	18423	31336	0.82	4.2E-01	BF242055.1	EST_HUMAN	601879721F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4108493 5'
5693	18487	31408	1.53	4.2E-01	AW854162.1	EST_HUMAN	RC3-CT0254-060400-029-g04 CT0254 Homo sapiens cDNA
6112	18889	31858	1.01	4.2E-01	AL163247.2	NT	Homo sapiens chromosome 21 segment HS21C047
6852	19552	32582	10.8	4.2E-01	AU158472.1	EST_HUMAN	AU158472 PLACE2 Homo sapiens cDNA clone PLACE2000470 3'
6852	19552	32583	10.8	4.2E-01	AU158472.1	EST_HUMAN	AU158472 PLACE2 Homo sapiens cDNA clone PLACE2000470 3'
6911	25101	32694	2.15	4.2E-01	S82504.1	NT	Brca1=breast cancer gene [rats, WF, spleen, Genomic, 419 nt, segment 2 of 2]
6993	19686	32734	7	4.2E-01	AL161547.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 47
7891	20586	33715	2.21	4.2E-01	AW957448.1	EST_HUMAN	EST369413 MAGE resequences, MAGE Homo sapiens cDNA
7891	20586	33716	2.21	4.2E-01	AW957448.1	EST_HUMAN	EST369413 MAGE resequences, MAGE Homo sapiens cDNA
8106	20800	33932	0.61	4.2E-01	4758039	NT	Homo sapiens cytochrome c oxidase subunit VIc (COX6C), nuclear gene encoding mitochondrial protein, mRNA
9870	22520		0.94	4.2E-01	AA705007.1	EST_HUMAN	zj95f01.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA clone IMAGE:462649 3'
10081	22729	35944	0.45	4.2E-01	AF181854.1	NT	Lassa virus strain 803213 glycoprotein precursor and nucleoprotein genes, complete cds
10393	23039	36256	1.78	4.2E-01	AW863666.1	EST_HUMAN	MR3-SN0010-280300-103-h07 SN0010 Homo sapiens cDNA
10972	23648	36901	2.69	4.2E-01	AB023489.1	NT	Oryzias latipes OlGC7 mRNA for membrane guanylyl cyclase, complete cds
11370	23977	37277	2.11	4.2E-01	BE966485.2	EST_HUMAN	601660352R1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906085 3'
1072	13830	26488	1.83	4.1E-01	AI905481.1	EST_HUMAN	RC-BT091-210199-142 BT091 Homo sapiens cDNA

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
1081	13839	26497	1.1	4.1E-01	AV705243.1	EST_HUMAN	AV705243 ADB Homo sapiens cDNA clone ADBAHF08 5'
1081	13839	26498	1.1	4.1E-01	AV705243.1	EST_HUMAN	AV705243 ADB Homo sapiens cDNA clone ADBAHF08 5'
2715	15422	28161	1.1	4.1E-01	7705283	NT	Homo sapiens anaphase-promoting complex subunit 7 (APC7), mRNA
2941	15706	28355	2.17	4.1E-01	AL161536.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 36
2941	15706	28356	2.17	4.1E-01	AL161536.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 36
3754	16506	29142	0.68	4.1E-01	AW961292.1	EST_HUMAN	EST373364 MAGE resequences, MAGG Homo sapiens cDNA
3754	16506	29143	0.68	4.1E-01	AW961292.1	EST_HUMAN	EST373364 MAGE resequences, MAGG Homo sapiens cDNA
4241	16982	29607	2.93	4.1E-01	AJ249207.1	NT	Rhodococcus sp. AD45 isoG, isoH, isoI, isoJ, isoA, isoB, isoC, isoD, isoE and isoF genes
4271	17011		0.82	4.1E-01	AA909257.1	EST_HUMAN	om33d02.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1542819 3'
4618	17353	29988	1.46	4.1E-01	AV747880.1	EST_HUMAN	AV747880 NPC Homo sapiens cDNA clone NPCBDF10 5'
4868	16057	28706	2.48	4.1E-01	AA906344.1	EST_HUMAN	oj94b08.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1505943 3'
5899	18684	31632	4.72	4.1E-01	BF681393.1	EST_HUMAN	602156590F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4297319 5'
7332	20014	33092	2.76	4.1E-01	U67535.1	NT	Methanococcus jannaschii section 77 of 150 of the complete genome
7935	20630	33757	1.38	4.1E-01	BF574604.1	EST_HUMAN	602133261F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4288238 5'
8988	21678	34827	1.39	4.1E-01	6755521	NT	Mus musculus signaling intermediate in Toll pathway-evolutionarily conserved (Sitpec-pending), mRNA
9465	22075		0.67	4.1E-01	AF160597.1	NT	Vcalavo gymnocaudus Vgym560 cytochrome b (cytb) gene, complete cds; mitochondrial gene for mitochondrial product
10163	22811		1.05	4.1E-01	AL139076.2	NT	Campylobacter jejuni NCTC11168 complete genome; segment 3/6
10310	22957	36173	0.91	4.1E-01	AV649579.1	EST_HUMAN	AV649579 GLC Homo sapiens cDNA clone GLCBVD12 3'
10404	23050	36267	0.61	4.1E-01	P18584	SWISSPROT	PROBABLE SERINE PROTEASE DO-LIKE PRECURSOR (59 KDA IMMUNOGENIC PROTEIN) (SK59)
10404	23050	36268	0.61	4.1E-01	P18584	SWISSPROT	PROBABLE SERINE PROTEASE DO-LIKE PRECURSOR (59 KDA IMMUNOGENIC PROTEIN) (SK59)
10478	23124		1.33	4.1E-01	BF349382.1	EST_HUMAN	CM2-HT0137-200999-010-e08 HT0137 Homo sapiens cDNA
10743	23430	36673	80.48	4.1E-01	X58700.1	NT	Zea mays ZMPMS2 gene for 19 kDa zein protein
11366	23177	36404	2	4.1E-01	Q09470	SWISSPROT	VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN KV1.1 (HUKI)(HBK1)
12475	25360		2.62	4.1E-01	D87675.1	NT	Homo sapiens DNA for amyloid precursor protein, complete cds
1016	13775	26435	0.82	4.0E-01	8404656	NT	Laqueus rubellus mitochondrion, complete genome
1316	14065	26739	0.95	4.0E-01	AF203478.1	NT	Drosophila melanogaster Dalmatian (dmt) mRNA, complete cds
1468	14215		4.05	4.0E-01	6679258	NT	Mus musculus platelet derived growth factor receptor, beta polypeptide (Pdgfrb), mRNA
1999	15583	27457	1.16	4.0E-01	Z96933.1	NT	Ascobolus immersus masc2 gene
1999	15583	27458	1.16	4.0E-01	Z96933.1	NT	Ascobolus immersus masc2 gene
2156	14886	27619	1.19	4.0E-01	AE001931.1	NT	Deinococcus radiodurans R1 section 68 of 229 of the complete chromosome 1
2156	14886	27620	1.19	4.0E-01	AE001931.1	NT	Deinococcus radiodurans R1 section 68 of 229 of the complete chromosome 1
2808	12952	25595	1.4	4.0E-01	6678490	NT	Mus musculus ubiquitin-protein ligase e3 componen n-recognin (Ubr1), mRNA

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
2968	15734	28383	1.1	4.0E-01	AL163280.2	NT	Homo sapiens chromosome 21 segment HS21C080
2968	15734	28384	1.1	4.0E-01	AL163280.2	NT	Homo sapiens chromosome 21 segment HS21C080
3683	16436	29080	1.98	4.0E-01	AF068903.1	NT	Streptococcus pneumoniae YllC (ylIC), YllD (ylID), penicillin-binding protein 2x (pbp2x), and undecaprenyl-phosphate-UDP-MurNAc-pentapeptide phospho-MurNAc-pentapeptide transferase (mraY) genes, complete cds
3807	16559	29191	3.38	4.0E-01	AJ277511.1	NT	Ovis aries partial JD2 gene for T cell receptor delta chain (TCRDJ2), exon 1
3807	16559	29192	3.38	4.0E-01	AJ277511.1	NT	Ovis aries partial JD2 gene for T cell receptor delta chain (TCRDJ2), exon 1
4767	17499		7.97	4.0E-01	Q31849	SWISSPROT	NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 5, CHLOROPLAST
5820	18609	31538	1.23	4.0E-01	AW970810.1	EST_HUMAN	EST382691 MAGE resequences, MAGK Homo sapiens cDNA
6345	19115	32104	0.94	4.0E-01	P27285	SWISSPROT	STRUCTURAL POLYPROTEIN (P130) [CONTAINS: COAT PROTEIN C ; SPIKE GLYCOPROTEINS E3, E2 AND E1; 6 KD PEPTIDE]
7728	20391	33504	0.66	4.0E-01	P27546	SWISSPROT	MICROTUBULE-ASSOCIATED PROTEIN 4
7829	20524	33649	0.44	4.0E-01	BF092634.1	EST_HUMAN	MR4-TN0110-180900-202-g02 TN0110 Homo sapiens cDNA
7910	20605	33736	1.04	4.0E-01	AB016625.1	NT	Homo sapiens OCTN2 gene, complete cds
8904	21595	34736	1.17	4.0E-01	AA323289.1	EST_HUMAN	EST26066 Cerebellum II Homo sapiens cDNA 5' end similar to EST containing Alu repeat
11560	24159		2.03	4.0E-01	BF030262.1	EST_HUMAN	601558283F1 NIH_MGC_58 Homo sapiens cDNA clone IMAGE:3828092 5'
11721	24315		2.83	4.0E-01	L76080.1	NT	Synechocystis sp. PCC 9413 transposase gene, complete cds
12162	25222		2.26	4.0E-01	AL163300.2	NT	Homo sapiens chromosome 21 segment HS21C100
12684	24972		2.2	4.0E-01	P36049	SWISSPROT	HYPOTHETICAL 49.7 KD PROTEIN IN GIN2-STE3 INTERGENIC REGION
1356	14104	26780	1.85	3.9E-01	AF206618.1	NT	Gorilla gorilla carboxyl-ester lipase (CEL) gene, complete cds
2648	15358	28101	3.34	3.9E-01	AB033019.1	NT	Homo sapiens mRNA for KIAA1193 protein, partial cds
2709	15416	28163	4.27	3.9E-01	X82032.1	NT	H.sapiens B-myb gene
2709	15416	28164	4.27	3.9E-01	X82032.1	NT	H.sapiens B-myb gene
3093	15858	28499	4.73	3.9E-01	AJ225896.1	NT	Sinorhizobium meliloti egl, syrB2, cya3 genes and orf3
4059	16804	29435	1.05	3.9E-01	BF592611.1	EST_HUMAN	7i61d01.x1 NCI_CGAP_Br16 Homo sapiens cDNA clone IMAGE:3339169 3'
4932	17660	30270	1.74	3.9E-01	BE728667.1	EST_HUMAN	601563948F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3833699 5'
5843	18631	31566	3.91	3.9E-01	BF208036.1	EST_HUMAN	601862362F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4082055 5'
7854	20549	33674	0.92	3.9E-01	U79415.1	NT	Homo sapiens prepro dipeptidyl peptidase I (DPP-I) gene, complete cds
8760	21452	34600	0.81	3.9E-01	AW177011.1	EST_HUMAN	CM3-CT0105-170899-004-b08 CT0105 Homo sapiens cDNA
8769	21461		0.58	3.9E-01	BF348634.1	EST_HUMAN	602019944F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4155322 5'
9134	21822	34988	1.26	3.9E-01	AW195888.1	EST_HUMAN	xn86d04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2701351 3' similar to TR:O94821 O94821 KIAA0713 PROTEIN ;
9445	22122	35301	1.46	3.9E-01	AI937337.1	EST_HUMAN	wp76a02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2467658 3' similar to SW:RFX5_HUMAN P48382 BINDING REGULATORY FACTOR. ;

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
9778	22429	35635	3.03	3.9E-01	M19879.1	NT	Human clabindin 27 gene, exons 10 and 11, and L1 and Alu repeats
9845	22496		0.58	3.9E-01	11465620	NT	Porphyra purpurea mitochondrion, complete genome
10066	22714	35932	0.77	3.9E-01	D86722.1	NT	Nicotiana tabacum mRNA for TATA binding protein (TBP), complete cds
10722	23410		1.98	3.9E-01	AV695974.1	EST_HUMAN	AV695974 GKC Homo sapiens cDNA clone GKCBQC11 5'
11753	24344	37674	1.47	3.9E-01	AV702623.1	EST_HUMAN	AV702623 ADB Homo sapiens cDNA clone ADBDBE06 5'
11948	25295		3.37	3.9E-01	AF304354.1	NT	Homo sapiens proteoglycan 3 (PRG3) gene, complete cds
12066	24581		2.08	3.9E-01	Q61670	SWISSPROT	HOMEOBOX PROTEIN HLX1
12559	24891		1.44	3.9E-01	11433335	NT	Homo sapiens hypothetical protein FLJ10583 (FLJ10583), mRNA
156	12971		8.33	3.8E-01	7019488	NT	Homo sapiens protein kinase PKNbeta (pknbeta), mRNA
1863	14601		1.03	3.8E-01	AE003870.1	NT	Xylella fastidiosa, section 16 of 229 of the complete genome
2460	15178	27918	1.29	3.8E-01	U41846.1	NT	Ceanorhabditis briggsae acetylcholinesterase (ace-1) gene, complete cds
2576	15290	28027	1.62	3.8E-01	AF214117.1	NT	Arabidopsis thaliana putative c-myb-like transcription factor (MYB3R-3) mRNA, complete cds
2638	15601	28092	3.96	3.8E-01	6678002	NT	Mus musculus solute carrier family 1, member 6 (Slc1a6), mRNA
3003	15769		1.14	3.8E-01	AJ251057.1	NT	Human immunodeficiency virus type 1 complete genome (isolate 98SE-MP1213)
3043	15809	28456	1.39	3.8E-01	AF043383.1	NT	Pleuronectes americanus aminopeptidase N (ampN) gene, partial cds
3477	16233	28887	7.98	3.8E-01	AL161518.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 30
3527	16283		0.79	3.8E-01	AI807219.1	EST_HUMAN	wf38b12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2357855 3'
3541	16283		1.22	3.8E-01	AI807219.1	EST_HUMAN	wf38b12.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2357855 3'
3739	16492	29127	1.16	3.8E-01	BE164080.1	EST_HUMAN	PM0-HT0339-200400-010-G01 HT0339 Homo sapiens cDNA
3897	16647	29287	0.97	3.8E-01	6754095	NT	Mus musculus general transcription factor I I I (Gtf2i), mRNA
4043	16788	29416	0.74	3.8E-01	AJ271361.2	NT	Takifugu rubripes wnt2 (partial), frank1, cftr and frank2 (partial) genes
5522	18320	31221	1.42	3.8E-01	Q04888	SWISSPROT	TRANSCRIPTION FACTOR SOX-10
6247	19021		0.74	3.8E-01	S46825.1	NT	prion protein [mink, Genomic, 2446 nt]
6528	19294	32298	5.6	3.8E-01	BE072399.1	EST_HUMAN	QV3-BT0537-271299-049-e02 BT0537 Homo sapiens cDNA
6862	19579	32614	4.58	3.8E-01	AI374601.1	EST_HUMAN	ta54f11.x1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone IMAGE:2047917 3' similar to contains Alu repetitive element;
6840	19502	32527	1.25	3.8E-01	AL161513.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 25
7416	20093		4.42	3.8E-01	X61597.1	NT	M.musculus gene for kallikrein-binding protein
8196	20890	34028	0.86	3.8E-01	M81385.1	NT	Mouse liver receptor homologous protein (LRH-1) mRNA, complete cds
8455	21147	34289	2.04	3.8E-01	AB046851.1	NT	Homo sapiens mRNA for KIAA1631 protein, partial cds
8523	21215	34358	1.02	3.8E-01	11441264	NT	Homo sapiens FOS-like antigen-1 (FOSL1), mRNA
8716	21408	34551	1.28	3.8E-01	AL163279.2	NT	Homo sapiens chromosome 21 segment HS21C079
9461	22011		3.55	3.8E-01	T95413.1	EST_HUMAN	ye43h06.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:120539 5' similar to contains Alu repetitive element;contains PTR5 repetitive element ;

Page 61 of 536
Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
10695	23386		1.67	3.8E-01	AV755814.1	EST_HUMAN	AV755814 BM Homo sapiens cDNA clone BMFBCE07 5'
11521	24121		3.18	3.8E-01	BE719219.1	EST_HUMAN	RC0-HT0841-040800-032-b12 HT0841 Homo sapiens cDNA
11693	24288	37610	2.27	3.8E-01	R42550.1	EST_HUMAN	yf92h11.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:30289 3'
11693	24288	37611	2.27	3.8E-01	R42550.1	EST_HUMAN	yf92h11.s1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:30289 3'
12149	24636		4.76	3.8E-01	AE001124.1	NT	Borrelia burgdorferi (section 10 of 70) of the complete genome
12270	25316		2.08	3.8E-01	U94788.1	NT	Human p53 (TP53) gene, complete cds
12384	24779		3.39	3.8E-01	BE829256.1	EST_HUMAN	QV3-ET0063-190700-271-a05 ET0063 Homo sapiens cDNA
12723	24994		1.54	3.8E-01	U78031.1	NT	Mus musculus apoptosis inhibitor bcl-x (bcl-x) gene, exon 3 and complete cds
12771	25291		1.74	3.8E-01	AF291483.1	NT	Mus musculus vomeronasal receptor V1RA4 (V1ra4) gene, complete cds
12788	25040	30966	1.51	3.8E-01	AF194972.1	NT	Mus musculus developmental control protein mRNA, partial cds
2486	15203	27944	12.24	3.7E-01	AB037831.1	NT	Homo sapiens mRNA for KIAA1410 protein, partial cds
3453	16209	28860	9.64	3.7E-01	AF056336.1	NT	Danio rerio bone morphogenetic protein 4 precursor (BMP4) gene, complete cds
4204	16945	29572	7.39	3.7E-01	AI218707.1	EST_HUMAN	ok39c07.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone IMAGE:1510188 3'
4286	17025	29651	1.3	3.7E-01	AW878037.1	EST_HUMAN	MR3-OT0007-080300-104-b02 OT0007 Homo sapiens cDNA
4357	17095	29730	2.55	3.7E-01	AE002408.1	NT	Neisseria meningitidis serogroup B strain MC58 section 50 of 206 of the complete genome
5676	18470	31386	1.15	3.7E-01	AF135187.1	NT	Homo sapiens interferon-induced protein p78 (MX1) gene, complete cds
5860	18647	31588	0.9	3.7E-01	AL163278.2	NT	Homo sapiens chromosome 21 segment HS21C078
6417	19185	32183	0.66	3.7E-01	M10806.1	NT	Chicken (White leghorn) delta-1 and delta-2 crystallin genes, complete cds
6436	19204		0.72	3.7E-01	L10353.1	NT	Mus saxicola haptoglobin mRNA, complete cds
7043	19734	32794	3.23	3.7E-01	11525843	NT	Homo sapiens tumor endothelial marker 7 precursor (TEM7), mRNA
7685	20349	33463	0.6	3.7E-01	T66802.1	EST_HUMAN	ya50a07.r3 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:66324 5'
7719	20383	33497	0.56	3.7E-01	AW511326.1	EST_HUMAN	hd45d05.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2912457 3' similar to contains Alu repetitive element;contains L1.t2 L1 repetitive element ;
8227	20921	34059	2.07	3.7E-01	11436739	NT	Homo sapiens chromosome 12 open reading frame 4 (C12ORF4), mRNA
8227	20921	34060	2.07	3.7E-01	11436739	NT	Homo sapiens chromosome 12 open reading frame 4 (C12ORF4), mRNA
8263	20957	34096	0.65	3.7E-01	AA902912.1	EST_HUMAN	ok43b11.s1 NCI_CGAP_Lei2 Homo sapiens cDNA clone IMAGE:1516701 3'
9101	21789		1.31	3.7E-01	AJ271386.1	NT	Gallus gallus mRNA for beta-carotene 15,15'-dioxygenase (bCDO gene)
10069	22717		0.8	3.7E-01	K00691.1	NT	mouse ig germline alpha membrane exons region
10110	22758	35970	4.12	3.7E-01	AI336411.1	EST_HUMAN	qt46b07.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:1950997 3'
10764	23448	36690	1.98	3.7E-01	X05958.1	NT	Rabbit mRNA for fast skeletal muscle myosin heavy chain (MHC)
10957	23633	36882	2.81	3.7E-01	AJ297357.1	NT	Homo sapiens partial LIMD1 gene for LIM domains containing protein 1 and KIAA0851 gene
10957	23633	36883	2.81	3.7E-01	AJ297357.1	NT	Homo sapiens partial LIMD1 gene for LIM domains containing protein 1 and KIAA0851 gene
11443	23210	36441	2.75	3.7E-01	X04122.1	NT	Bovine mRNA for terminal deoxynucleotidyltransferase (TdT) (EC 2.7.7.31)
11676	24271	37593	1.43	3.7E-01	D79348.1	EST_HUMAN	HUM230A06B Human aorta polyA+ (TFujiwara) Homo sapiens cDNA clone GEN-230A06 5'

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
11771	24362		2.87	3.7E-01	6677678	NT	Mus musculus retinoblastoma 1 (Rb1), mRNA
11869	24943		2.11	3.7E-01	J04982.1	NT	Human heart/skeletal muscle ATP/ADP translocator (ANT1) gene, complete cds
12033	24558		3.09	3.7E-01	AJ243525.1	NT	Chlamydophila psittaci partial omp1 gene for outer membrane protein 1
12488	24847		1.9	3.7E-01	AL121154.1	EST_HUMAN	DKFZp762K075_r1 762 (synonym: hmel2) Homo sapiens cDNA clone DKFZp762K075 5'
12548	24886	30995	4.03	3.7E-01	Y18000.1	NT	Homo sapiens NF2 gene
254	13062	25701	2.17	3.6E-01	AJ009609.1	NT	Brassica napus mRNA for MAP4K alpha2 protein
975	13740		8.22	3.6E-01	U89241.1	NT	Human mlbp gene, partial cds
1291	14040	26713	3.83	3.6E-01	T80255.1	EST_HUMAN	yd03e05.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:24443 5'
1291	14040	26714	3.83	3.6E-01	T80255.1	EST_HUMAN	yd03e05.r1 Soares infant brain 1NIB Homo sapiens cDNA clone IMAGE:24443 5'
1909	14646	27356	6.73	3.6E-01	AW590184.1	EST_HUMAN	hg33f02.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2947419 3'
1909	14646	27357	6.73	3.6E-01	AW590184.1	EST_HUMAN	hg33f02.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2947419 3'
1944	14679	27393	5.7	3.6E-01	AF216207.1	NT	Mus musculus ribosomal protein S19 (Rps19) gene, complete cds
2047	14780		1.39	3.6E-01	AF056927.1	NT	Rattus norvegicus repeat element associated with the Rasgrf1 gene
2267	14993		1.05	3.6E-01	AB002321.1	NT	Human mRNA for KIAA0323 gene, partial cds
2389	15110		2.69	3.6E-01	X76725.1	NT	P.irregulare (P3804) gene for actin
2479	15197	27936	1.23	3.6E-01	L05435.1	NT	Rattus norvegicus synaptic vesicle protein (SV2) mRNA, complete cds
2479	15197	27937	1.23	3.6E-01	L05435.1	NT	Rattus norvegicus synaptic vesicle protein (SV2) mRNA, complete cds
2491	15208	27950	1.43	3.6E-01	AW812033.1	EST_HUMAN	RC5-ST0171-181099-011-g07 ST0171 Homo sapiens cDNA
2636	15348	28090	1.44	3.6E-01	P24206	SWISSPROT	PROTEIN-L-ISOASPARTATE O-METHYLTRANSFERASE (PROTEIN-BETA-ASPARTATE METHYLTRANSFERASE) (PIMT) (PROTEIN L-ISOASPARTYL METHYLTRANSFERASE) (L-ISOASPARTYL PROTEIN CARBOXYL METHYLTRANSFERASE)
2900	17884		7.16	3.6E-01	AF199485.1	NT	Drosophila melanogaster sugar transporter 3 (sut3) mRNA, complete cds
3462	16218	28871	2.16	3.6E-01	X76758.1	NT	H.sapiens serotonin transporter gene, exons 9 and 10
3462	16218	28872	2.16	3.6E-01	X76758.1	NT	H.sapiens serotonin transporter gene, exons 9 and 10
4375	17112	29745	1.3	3.6E-01	BE707883.1	EST_HUMAN	RC1-HT0545-150600-014-b12 HT0545 Homo sapiens cDNA
4948	17675	30285	2.38	3.6E-01	AW339393.1	EST_HUMAN	ha02g04.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2872566 3'
5298	18103	30762	0.82	3.6E-01	AJ006565.1	NT	Homo sapiens lipe gene intron 5
5995	18776	31738	0.85	3.6E-01	P16431	SWISSPROT	FORMATE HYDROGENLYASE SUBUNIT 5 PRECURSOR (FHL SUBUNIT 5) (HYDROGENASE-3 COMPONENT E)
6386	19155	32154	1.74	3.6E-01	Y10196.1	NT	Homo sapiens PHEX gene
7048	19739		3.2	3.6E-01	R94090.1	EST_HUMAN	yt74a06.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:275987 5'
7183	19869	32943	1.9	3.6E-01	AW027174.1	EST_HUMAN	wt72c10.x1 Soares_thymus_NHFTh Homo sapiens cDNA clone IMAGE:2513010 3' similar to TR:O15117 O15117 FYN BINDING PROTEIN. [1];
8123	20817	33953	0.58	3.6E-01	P98167	SWISSPROT	SCO-SPONDIN

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8177	20871	34005	11.45	3.6E-01	AL161583.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 79
8900	21591	34731	2.74	3.6E-01	4504958	NT	Homo sapiens lysosomal-associated membrane protein 2 (LAMP2), transcript variant LAMP2A, mRNA
8900	21591	34732	2.74	3.6E-01	4504958	NT	Homo sapiens lysosomal-associated membrane protein 2 (LAMP2), transcript variant LAMP2A, mRNA
9091	21780	34944	1.17	3.6E-01	AL163204.2	NT	Homo sapiens chromosome 21 segment HS21C004
9299	21966	35139	1.04	3.6E-01	X17550.1	NT	D. melanogaster singed gene, exons 3, 4, 5 & 6
9299	21966	35140	1.04	3.6E-01	X17550.1	NT	D. melanogaster singed gene, exons 3, 4, 5 & 6
9369	21944		0.57	3.6E-01	X62825.1	NT	C.perfringens plc gene for phospholipase C upstream region containing bent DNA fragment
9763	22414	35621	14.67	3.6E-01	Q53194	SWISSPROT	PROBABLE PEPTIDE ABC TRANSPORTER ATP-BINDING PROTEIN Y4TS
9893	22543	35735	0.51	3.6E-01	AW752901.1	EST_HUMAN	MR2-CT0222-211099-002-b10 CT0222 Homo sapiens cDNA
9893	22543	35736	0.51	3.6E-01	AW752901.1	EST_HUMAN	MR2-CT0222-211099-002-b10 CT0222 Homo sapiens cDNA
10864	23544	36791	3.31	3.6E-01	BE902390.1	EST_HUMAN	601676418F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3958997 5'
11052	23722	36993	4.12	3.6E-01	AB004293.1	NT	Arabidopsis thaliana mRNA for SigB, complete cds
11421	23188	36419	3.4	3.6E-01	AE000856.1	NT	Methanobacterium thermoautotrophicum from bases 702375 to 714311 (section 62 of 148) of the complete genome
11903	25416		1.83	3.6E-01	Y19210.1	NT	Homo sapiens hHb5 gene for hair keratin, exons 1 to 9
11978	24522		1.4	3.6E-01	D90901.1	NT	Synechocystis sp. PCC6803 complete genome, 3/27, 271600-402289
11987	24528		3.89	3.6E-01	AE000335.1	NT	Escherichia coli K-12 MG1655 section 225 of 400 of the complete genome
12135	24624		4	3.6E-01	U66888.1	NT	Mus musculus Emr1 mRNA, complete cds
12493	24850		2.12	3.6E-01	11432598	NT	Homo sapiens myeloid/lymphoid or mixed-lineage leukemia (trithorax (Drosophila) homolog); translocated to, 10 (AF10), mRNA
12746	25363		2.23	3.6E-01	AW190229.1	EST_HUMAN	xl60e11.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2679116 3' similar to gb:K00558 TUBULIN ALPHA-1 CHAIN (HUMAN);
204	13017	25657	2.05	3.5E-01	6678933	NT	Mus musculus mannose receptor, C type 2 (Mrc2), mRNA
708	13482	26131	1.59	3.5E-01	7706136	NT	Homo sapiens GAP-like protein (LOC51306), mRNA
708	13482	26132	1.59	3.5E-01	7706136	NT	Homo sapiens GAP-like protein (LOC51306), mRNA
762	13535	26194	4.25	3.5E-01	BF129796.1	EST_HUMAN	601811060R1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4053951 3'
1615	14362	27053	1.1	3.5E-01	BF310688.1	EST_HUMAN	601894653F2 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4124244 5'
1636	14382	27069	1.96	3.5E-01	U35776.1	NT	Rattus norvegicus ADP-ribosylation factor-directed GTPase activating protein mRNA, complete cds
2281	15006	27747	1.35	3.5E-01	P06798	SWISSPROT	HOMEOBOX PROTEIN HOX-A4 (HOX-1.4) (MH-3)
2812	15600	28066	1.76	3.5E-01	AA223252.1	EST_HUMAN	zr08a09.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens cDNA clone IMAGE:650872 3'

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
3785	16537		0.85	3.5E-01	AA642138.1	EST_HUMAN	nr60d03.s1 NCI_CGAP_Lym3 Homo sapiens cDNA clone IMAGE:1172357 3'
4231	16972	29596	1.67	3.5E-01	AF071253.1	NT	Danio rerio homeobox protein (hoxb5b) gene, complete cds
4443	17179	29805	0.94	3.5E-01	BE146585.1	EST_HUMAN	RC5-HT0218-181099-011-g02 HT0218 Homo sapiens cDNA
4627	17362	29995	1.02	3.5E-01	Y18477.1	NT	Mus musculus Alox12B gene 5' flanking region
4880	17607	30230	4.58	3.5E-01	M18349.1	NT	Rat leukocyte common antigen (L-CA) gene, exons 1 through 5
5251	18057	30685	0.76	3.5E-01	Q96687	SWISSPROT	EARLY E2A DNA-BINDING PROTEIN
5251	18057	30686	0.76	3.5E-01	Q96687	SWISSPROT	EARLY E2A DNA-BINDING PROTEIN
5462	18261	31152	1.13	3.5E-01	D42045.1	NT	Human mRNA for KIAA0086 gene, complete cds
6143	18921		0.96	3.5E-01	AW863916.1	EST_HUMAN	PM4-SN0012-030400-001-a11 SN0012 Homo sapiens cDNA
6314	19085	32070	0.6	3.5E-01	AA431833.1	EST_HUMAN	zw79f03.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:782429 5' similar to TR:G1066935 G1066935 F10F2.1 ;
6359	19129	32124	0.66	3.5E-01	U37150.1	NT	Bos taurus peptide methionine sulfoxide reductase (msrA) mRNA, complete cds
6566	19331	32338	1.08	3.5E-01	O24357	SWISSPROT	GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE, CHLOROPLAST PRECURSOR (G6PD)
6956	19438		4.24	3.5E-01	X98505.1	NT	S.scrofa mRNA for CD31 protein (PECAM-1)
7441	20118	33207	0.55	3.5E-01	P47281	SWISSPROT	HISTIDYL-TRNA SYNTHETASE (HISTIDINE--TRNA LIGASE) (HISRS)
7441	20118	33208	0.55	3.5E-01	P47281	SWISSPROT	HISTIDYL-TRNA SYNTHETASE (HISTIDINE--TRNA LIGASE) (HISRS)
7970	20665		2.19	3.5E-01	11448042	NT	Homo sapiens tumor protein p53-binding protein, 2 (TP53BP2), mRNA
7973	20668	33790	0.71	3.5E-01	BF358871.1	EST_HUMAN	RC4-ET0024-260600-014-d07 ET0024 Homo sapiens cDNA
8366	21059		0.63	3.5E-01	AF051561.1	NT	Rattus norvegicus Na-K-Cl cotransporter (Nkcc1) mRNA, complete cds
8825	21517	34662	1.17	3.5E-01	4507610	NT	Homo sapiens tyrosine kinase non-receceptor 1 (TNK1), mRNA
9636	22288	35481	1.52	3.5E-01	Q02294	SWISSPROT	VOLTAGE-DEPENDENT N-TYPE CALCIUM CHANNEL ALPHA-1B SUBUNIT (CALCIUM CHANNEL, L TYPE, ALPHA-1 POLYPEPTIDE ISOFORM 5) (BRAIN CALCIUM CHANNEL III) (BIII)
9786	22437	35644	5.64	3.5E-01	Z26825.1	NT	X.laevis gene for albumin including HP1 enhancer
9867	22517	35713	0.96	3.5E-01	BE174794.1	EST_HUMAN	QV2-HT0577-090400-128-c07 HT0577 Homo sapiens cDNA
10635	23327	36564	2.76	3.5E-01	X61084.1	NT	C.griseus rhodopsin gene for opsin protein
10946	23625	36875	2.39	3.5E-01	AJ243178.1	NT	Gallus gallus SPARC gene for osteonectin, promoter and exon 1
10946	23625	36876	2.39	3.5E-01	AJ243178.1	NT	Gallus gallus SPARC gene for osteonectin, promoter and exon 1
11505	24106	37419	1.34	3.5E-01	U07000.1	NT	Human breakpoint cluster region (BCR) gene, complete cds
11585	24184	37499	1.64	3.5E-01	N77597.1	EST_HUMAN	yz90h12.r1 Soares_multiple_sclerosis_2NbHMSP Homo sapiens cDNA clone IMAGE:290375 5'
11619	24216		1.71	3.5E-01	M82885.1	NT	Drosophila melanogaster dual bar protein (BarH2) gene, exon 1
11684	24279	37601	1.51	3.5E-01	L05145.1	NT	Human glucokinase (GCK) gene, repeat polymorphism
11776	24367		1.36	3.5E-01	AI064773.1	EST_HUMAN	HA0542 Human fetal liver cDNA library Homo sapiens cDNA
12063	24578		1.47	3.5E-01	X64565.1	NT	B.taurus atpA1 gene for F(0)F(1) ATP synthase alpha-subunit
12214	24876		2.32	3.5E-01	AE001774.1	NT	Thermotoga maritima section 86 of 136 of the complete genome

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
12402	24787		1.4	3.5E-01	AE001691.1	NT	Thermotoga maritima section 3 of 136 of the complete genome
12793	25269	30723	3.33	3.5E-01	H80814.1	EST_HUMAN	ys64f11.r1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:219597 5'
12793	25269	30724	3.33	3.5E-01	H80814.1	EST_HUMAN	ys64f11.r1 Soares retina N2b4HR Homo sapiens cDNA clone IMAGE:219597 5'
691	13466		1.85	3.4E-01	AJ242956.1	NT	Homo sapiens partial N-myc (exon 3), HPV45 L2, HPV45 L1, HPV45 E6, HPV45 E7 and HPV45 E1 genes isolated from IC4 cervical carcinoma cell line
955	13720	26386	7.61	3.4E-01	Y09798.2	NT	Pseudomonas fluorescens colR, colS genes, orf222 and partial inaA gene
1303	14052	26725	1.72	3.4E-01	Y00554.1	NT	Azotobacter vinelandii nifA gene for NifA protein (positive regulatory element)
2400	15121	27858	2.62	3.4E-01	D90909.1	NT	Synechocystis sp. PCC6803 complete genome, 11/27, 1311235-1430418
3001	15767	28415	0.85	3.4E-01	AL163210.2	NT	Homo sapiens chromosome 21 segment HS21C010
3001	15767	28416	0.85	3.4E-01	AL163210.2	NT	Homo sapiens chromosome 21 segment HS21C010
3146	15910	28555	1.08	3.4E-01	D90909.1	NT	Synechocystis sp. PCC6803 complete genome, 11/27, 1311235-1430418
3159	15922	28568	6.23	3.4E-01	U83905.1	NT	Canis familiaris rod photoreceptor cGMP-gated channel alpha-subunit (CNGC1) mRNA, complete cds
3338	16098	28749	0.9	3.4E-01	AF034862.1	NT	Homo sapiens pulmonary surfactant protein D, promoter region and exon 1
3522	16278	28933	3.48	3.4E-01	AF106835.1	NT	Methylovorus sp. strain SS1 putative GrpE (grpE), DnaK (dnaK), and putative DnaJ (dnaJ) genes, complete cds
3770	16522		1.69	3.4E-01	BF449010.1	EST_HUMAN	7n94a01.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3572232 3' similar to TR:Q9UJ15 Q9UJ15 DJ18C9.1 ;
4029	16774		2.38	3.4E-01	AA584198.1	EST_HUMAN	no11b10.s1 NCI_CGAP_Phe1 Homo sapiens cDNA clone IMAGE:1100347 3'
4460	17196	29823	0.82	3.4E-01	AF166341.1	NT	Homo sapiens integrin alpha 6 (ITGA6) gene, exons 12 through 23
4599	17334	29963	1.54	3.4E-01	BE069912.1	EST_HUMAN	MR4-BT0403-230200-202-c01 BT0403 Homo sapiens cDNA
4898	17625		3.23	3.4E-01	AI240973.1	EST_HUMAN	qj95c05.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1867208 3' similar to contains Alu repetitive element;
5143	17862		0.98	3.4E-01	U79746.1	NT	Homo sapiens serotonin transporter (hSERT) gene, promoter region, exons 1B and 2, and partial cds
5599	18394	31304	2.62	3.4E-01	AL161594.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 90
5721	18513		6.09	3.4E-01	AA085313.1	EST_HUMAN	zn12d11.s1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone IMAGE:547221 3'
5917	18702		1.99	3.4E-01	L02971.1	NT	Echovirus 22 1AB, 1C, 1D, 2A, 2B, 2C, 3A, 3B, 3C, 3D proteins RNA, complete mature peptides and cds
5940	18722	31681	0.89	3.4E-01	BE748912.1	EST_HUMAN	601571811T1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:3838826 3'
6017	18798	31759	2.43	3.4E-01	AW204505.1	EST_HUMAN	UI-H-BI1-asi-e-12-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone IMAGE:2719582 3'
6141	18919	31889	1.81	3.4E-01	AL120544.1	EST_HUMAN	DKFZp761A249_r1 761 (synonym: hamy2) Homo sapiens cDNA clone DKFZp761A249 5'
6644	19406		1.56	3.4E-01	N95225.1	EST_HUMAN	zb53e12.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:307342 3'

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
6848	19548	32578	1.02	3.4E-01	AI468082.1	EST_HUMAN	tm63g05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2162840 3' similar to gb:S37431 LAMININ RECEPTOR (HUMAN);
6959	19441	32456	0.59	3.4E-01	BF678702.1	EST_HUMAN	602085283F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4249365 5'
7806	20501		0.49	3.4E-01	AE000493.1	NT	Escherichia coli K-12 MG1655 section 383 of 400 of the complete genome
8135	20829	33964	0.6	3.4E-01	Y14930.1	NT	Homo sapiens TCRAV28 gene, allele A4, partial
8188	20882		0.47	3.4E-01	BF449010.1	EST_HUMAN	7n94a01.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3572232 3' similar to TR:Q9UJ15 Q9UJ15 DJ18C9.1 ;
8386	21079		1.51	3.4E-01	AA337063.1	EST_HUMAN	EST41765 Endometrial tumor Homo sapiens cDNA 5' end
8461	21153	34296	0.72	3.4E-01	L04690.1	NT	Cricetulus griseus cholesterol 7-alpha-hydroxylase gene, complete cds
8751	21443	34590	1.7	3.4E-01	9633624	NT	Bovine enterovirus strain K2577, complete genome
9112	21800	34964	4.42	3.4E-01	P26013	SWISSPROT	INTEGRIN BETA-8 PRECURSOR
9112	21800	34965	4.42	3.4E-01	P26013	SWISSPROT	INTEGRIN BETA-8 PRECURSOR
9321	21988		0.51	3.4E-01	AB017510.1	NT	Ephydatia fluviatilis mRNA for PLC-gammaS, complete cds
9346	20417	33536	4.67	3.4E-01	U19492.1	NT	Saccharomyces cerevisiae Maf1p (MAF1) gene, complete cds
9346	20417	33537	4.67	3.4E-01	U19492.1	NT	Saccharomyces cerevisiae Maf1p (MAF1) gene, complete cds
9397	22059	35229	0.5	3.4E-01	AF193857.1	NT	Dictyostelium discoideum putative CMF receptor CMFR1 mRNA, complete cds
9595	22248	35433	1.01	3.4E-01	U68763.1	NT	Glycine max putative transcription factor SCOF-1 (scof-1) mRNA, complete cds
9789	22440	35648	1.86	3.4E-01	AJ225084.1	NT	Homo sapiens FAA gene, exon 16, 17 and 18
10376	23022		0.62	3.4E-01	AE004096.1	NT	Vibrio cholerae chromosome I, section 4 of 251 of the complete chromosome
10940	23620		4.72	3.4E-01	AE000881.1	NT	Methanobacterium thermoautotrophicum from bases 1018444 to 1029212 (section 87 of 148) of the complete genome
10984	23659	36912	2.6	3.4E-01	P06925	SWISSPROT	PROBABLE E4 PROTEIN
11032	23703	36971	2.17	3.4E-01	AF045981.1	NT	Rutilus arcasii cytochrome b (cytb) gene, mitochondrial gene encoding mitochondrial protein, partial cds
11253	23915	37207	1.61	3.4E-01	M25856.1	NT	Human von Willebrand factor gene, exons 36 and 37
11253	23915	37208	1.61	3.4E-01	M25856.1	NT	Human von Willebrand factor gene, exons 36 and 37
11483	24084	37396	1.88	3.4E-01	AB035507.1	NT	Rattus norvegicus mRNA for s-gicerin/MUC18, complete cds
11513	24113	37423	3.65	3.4E-01	AL161515.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 27
11786	24376	37706	1.72	3.4E-01	BF061948.1	EST_HUMAN	7k69d12.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:3480646 3'
11861	24445	37786	1.58	3.4E-01	U07000.1	NT	Human breakpoint cluster region (BCR) gene, complete cds
11881	24455		1.85	3.4E-01	U93604.1	NT	Citrus variegation virus putative replicase gene, partial cds
12197	24666		11.43	3.4E-01	L26339.1	NT	Human autoantigen mRNA, complete cds
12224	25192		1.61	3.4E-01	BE218652.1	EST_HUMAN	hv42h08.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3176127 3' similar to contains PTR5.t3 PTR5 repetitive element ;

Table 4

Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
12280	25292		2.28	3.4E-01	9838361	NT	Beta vulgaris mitochondrion, complete genome
12391	24781	31036	2.2	3.4E-01	AJ297131.1	NT	Mus musculus SIL, MAP_17, CYP_a, SCL & CYP_b genes
12688	24974		1.82	3.4E-01	AF019413.1	NT	Homo sapiens HLA class III region containing tenascin X (tenascin-X) gene, partial cds; cytochrome P450 21-hydroxylase (CYP21B), complement component C4 (C4B) G11, helicase (SKI2W), RD, complement factor B (Bf), and complement component C2 (C2) genes,>
13	12840	25453	10.77	3.3E-01	X07990.1	NT	Rhizobium leguminosarum sym plasmid pRL5JI nodX gene
103	12840	25453	4.4	3.3E-01	X07990.1	NT	Rhizobium leguminosarum sym plasmid pRL5JI nodX gene
435	13221	25867	0.9	3.3E-01	AL161545.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 45
618	13397	26032	2.01	3.3E-01	7662485	NT	Homo sapiens KIAA1100 protein (KIAA1100), mRNA
1178	13931	26597	2.85	3.3E-01	Q12446	SWISSPROT	PROLINE-RICH PROTEIN LAS17
1284	14034	26705	3.76	3.3E-01	BF568880.1	EST_HUMAN	602184016T1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4300251 3'
1336	14085	26760	1.2	3.3E-01	U43628.1	NT	Human chromosome 15q11-q13 putative DNA replication origin in the g-aminobutyric acid receptor b3 and a5 gene cluster
1601	14347	27036	1.47	3.3E-01	6753685	NT	Mus musculus disintegrin 5 (Dtgn5), mRNA
1731	14473		1.02	3.3E-01	AA332734.1	EST_HUMAN	EST36722 Embryo, 8 week I Homo sapiens cDNA 5' end
2022	14757		1.01	3.3E-01	AF031148.1	NT	Methylococcus capsulatus strain Bath outer membrane protein MopB (mopB) gene, complete cds
2404	15125		4.62	3.3E-01	4507834	NT	Homo sapiens uridine monophosphate synthetase (orotate phosphoribosyl transferase and orotidine-5'-decarboxylase) (UMPS) mRNA
2949	15715	28368	1.87	3.3E-01	AJ251805.1	NT	Bacteriophage phi-YeO3-12 complete genome
3051	15817	28462	1.48	3.3E-01	AJ007932.2	NT	Streptomyces argillaceus mithramycin biosynthetic genes
3486	16243	28899	1.07	3.3E-01	AB012922.1	NT	Homo sapiens MTA1-L1 gene, complete cds
3789	16541	29176	2.1	3.3E-01	O84645	SWISSPROT	EXODEOXYRIBONUCLEASE V BETA CHAIN
3799	16551	29183	0.97	3.3E-01	P22602	SWISSPROT	GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (P1); HELPER COMPONENT PROTEINASE (HC-PRO); PROTEIN P3]
3932	16682	29323	1.03	3.3E-01	4757739	NT	Homo sapiens A kinase (PRKA) anchor protein 5 (AKAP5), mRNA
3947	16697	29336	1.47	3.3E-01	AL161498.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 10
3983	16731	29365	1.79	3.3E-01	AF200446.1	NT	Hypoxylon fragiforme chitin synthase gene, partial cds
4334	17073		1.6	3.3E-01	D31662.1	NT	Rattus norvegicus DNA for regucalcin, partial cds
4641	17375		1.23	3.3E-01	AI539114.1	EST_HUMAN	tp78b12.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2205407 3' similar to gb:X57522 ANTIGEN PEPTIDE TRANSPORTER 1 (HUMAN);
4786	17517	30139	1.22	3.3E-01	D64003.1	NT	Synechocystis sp. PCC6803 complete genome, 22/27, 2755703-2868766
5146	17865		0.96	3.3E-01	AW937982.1	EST_HUMAN	QV0-DT0047-170200-123-h08 DT0047 Homo sapiens cDNA
5241	18047	30675	2.61	3.3E-01	X89819.1	NT	R.norvegicus mRNA for 3'UTR of ubiquitin-like protein
5241	18047	30676	2.61	3.3E-01	X89819.1	NT	R.norvegicus mRNA for 3'UTR of ubiquitin-like protein

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
5700	18494	31417	0.74	3.3E-01	BF213873.1	EST_HUMAN	601848090F1 NIH_MGC_55 Homo sapiens cDNA clone IMAGE:4078823 5'
5856	18643	31582	1.9	3.3E-01	BE619650.1	EST_HUMAN	601472768T1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3875753 3'
5856	18643	31583	1.9	3.3E-01	BE619650.1	EST_HUMAN	601472768T1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3875753 3'
5947	18729	31688	1.18	3.3E-01	P05691	SWISSPROT	CIRCUMSPOROZOITE PROTEIN (CS)
6695	19612	32651	0.71	3.3E-01	AB034233.1	NT	Flexibacter litoralis gyrB gene for DNA gyrase B subunit, partial cds
6695	19612	32652	0.71	3.3E-01	AB034233.1	NT	Flexibacter litoralis gyrB gene for DNA gyrase B subunit, partial cds
6789	19533	32560	4.82	3.3E-01	AI628131.1	EST_HUMAN	ty64h01.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2285809 3' similar to contains Alu repetitive element;contains element L1 repetitive element ;
6789	19533	32561	4.82	3.3E-01	AI628131.1	EST_HUMAN	ty64h01.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2285809 3' similar to contains Alu repetitive element;contains element L1 repetitive element ;
7682	20346	33458	1.68	3.3E-01	N85146.1	EST_HUMAN	J2498F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA clone J2498 5' similar to TEGT
8460	21152	34295	18.62	3.3E-01	BF683954.1	EST_HUMAN	602140372F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4301800 5'
8659	21351	34497	0.48	3.3E-01	AU126115.1	EST_HUMAN	AU126115 NT2RP1 Homo sapiens cDNA clone NT2RP1000130 5'
8659	21351	34498	0.48	3.3E-01	AU126115.1	EST_HUMAN	AU126115 NT2RP1 Homo sapiens cDNA clone NT2RP1000130 5'
9012	21702	34852	0.81	3.3E-01	Q62925	SWISSPROT	MITOGEN-ACTIVATED PROTEIN KINASE KINASE KINASE 1 (MAPK/ERK KINASE KINASE 1) (MEK KINASE 1) (MEKK 1)
9278	22032	35203	0.81	3.3E-01	BE828461.1	EST_HUMAN	CM3-ET0041-180500-187-d10 ET0041 Homo sapiens cDNA
9278	22032	35204	0.81	3.3E-01	BE828461.1	EST_HUMAN	CM3-ET0041-180500-187-d10 ET0041 Homo sapiens cDNA
9411	22073	35244	2.62	3.3E-01	N69866.1	EST_HUMAN	za67h01.s1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone IMAGE:297649 3'
9452	22002	35174	2.77	3.3E-01	BF376745.1	EST_HUMAN	RC4-TN0077-250800-011-g04 TN0077 Homo sapiens cDNA
9891	22541		2.27	3.3E-01	L41044.1	NT	Homo sapiens high-mobility group phosphoprotein (HMGI-C) gene, exons 1-3, complete cds
10622	23315	36554	3.13	3.3E-01	X63953.1	NT	D.mauritiana Adh gene
10622	23315	36555	3.13	3.3E-01	X63953.1	NT	D.mauritiana Adh gene
10951	23628		1.7	3.3E-01	BF526499.1	EST_HUMAN	602070802F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4213585 5'
11196	23861	37147	11.61	3.3E-01	BE219351.1	EST_HUMAN	hv51g02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3176978 3'
11317	24008	37313	3.23	3.3E-01	P47953	SWISSPROT	GALECTIN-3 (GALACTOSE-SPECIFIC LECTIN 3) (MAC-2 ANTIGEN) (IGE-BINDING PROTEIN) (35 KD LECTIN) (CARBOHYDRATE BINDING PROTEIN 35) (CBP 35) (LAMININ-BINDING PROTEIN) (LECTIN L-29) (CBP30)
11719	24313		3.06	3.3E-01	AA806621.1	EST_HUMAN	ob71g02.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336850 3'
11741	12840	25453	1.87	3.3E-01	X07990.1	NT	Rhizobium leguminosarum sym plasmid pRL5JI nodX gene
11977	24521	37266	1.71	3.3E-01	6598319	NT	Homo sapiens aldehyde oxidase 1 (AOX1), mRNA
12676	24967		3.34	3.3E-01	AP000002.1	NT	Pyrococcus horikoshii OT3 genomic DNA, 287001-544000 nt. position (2/7)
444	13230		2.33	3.2E-01	AF018261.1	NT	Rattus norvegicus EH domain binding protein Epsin mRNA, complete cds

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
701	13476		1.43	3.2E-01	AL161561.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 61
1139	13894	26555	27.53	3.2E-01	AF047013.1	NT	Fusarium poae virus 1 RNA2 putative RNA dependent RNA polymerase gene, complete cds
1259	14008	26677	1.36	3.2E-01	Z50202.1	NT	P.vulgaris arc5-1 gene
1369	14117	26792	5.42	3.2E-01	Q48624	SWISSPROT	LACTOSE PERMEASE (LACTOSE-PROTON SYMPORT) (LACTOSE TRANSPORT PROTEIN)
1767	14509	27210	1.25	3.2E-01	Z36041.1	NT	S.cerevisiae chromosome II reading frame ORF YBR172c
1777	14519	27222	4.7	3.2E-01	AW957194.1	EST_HUMAN	EST369264 MAGE resequences, MAGD Homo sapiens cDNA
1777	14519	27223	4.7	3.2E-01	AW957194.1	EST_HUMAN	EST369264 MAGE resequences, MAGD Homo sapiens cDNA
1835	14574	27286	1.23	3.2E-01	AL111655.1	NT	Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation
2157	14887	27621	2.52	3.2E-01	BF203817.1	EST_HUMAN	601868804F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4111512 5'
2543	15257		2.01	3.2E-01	7710079	NT	Mus musculus Pbx/knotted 1 homeobox (Pknox1), mRNA
2713	15420	28159	1.08	3.2E-01	AF060568.1	NT	Homo sapiens promyelocytic leukemia zinc finger protein (PLZF) gene, complete cds
3594	16347		0.77	3.2E-01	D10872.1	NT	Humam h NAT allele 3-2 gene for arylamine N-acetyltransferase
4305	17044	29669	0.91	3.2E-01	4759195	NT	Homo sapiens symplekin (SYM) mRNA
4363	17101	29736	1.52	3.2E-01	M18818.1	NT	Rabbit beta-like globin gene cluster encoding the epsilon, gamma, delta (pseudogene) and beta globin polypeptides, complete cds
4464	17200	29826	1.21	3.2E-01	Q10268	SWISSPROT	HYPOTHETICAL 81.7 KD PROTEIN C13G7.04C IN CHROMOSOME I PRECURSOR
4688	17422		6.7	3.2E-01	BF693617.1	EST_HUMAN	602081972F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4246505 5'
4826	17557	30179	1.17	3.2E-01	Q57081	SWISSPROT	CYTADHERENCE HIGH MOLECULAR WEIGHT PROTEIN 3 (CYTADHERENCE ACCESSORY PROTEIN 3) (ACCESSORY ADHESIN PROTEIN 3) (P69)
4965	17690	30299	0.74	3.2E-01	BE782748.1	EST_HUMAN	601465591F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3868799 5'
5190	17998	30621	3.26	3.2E-01	BE173964.1	EST_HUMAN	CM0-HT0569-060300-269-f10 HT0569 Homo sapiens cDNA
5868	18655	31596	1.07	3.2E-01	L27221.1	NT	Giardia intestinalis pyruvate:flavodoxin oxidoreductase and flanking genes
6211	18986	31963	0.9	3.2E-01	AF016494.1	NT	Fugu rubripes gamma-aminobutyric acid receptor beta subunit gene, partial cds; 55kd erythrocyte membrane protein (P55), synaptic vesicle-associated integral membrane protein (VAMP-1), procollagen C-proteinase enhancer protein (PCOLCE) genes, complete c>
6501	19266	32268	0.64	3.2E-01	AV718037.1	EST_HUMAN	AV718037 FHTA Homo sapiens cDNA clone FHTAABH01 5'
6634	19396		1.09	3.2E-01	AB002359.1	NT	Human mRNA for KIAA0361 gene, KIAA0361 protein
7755	20451	33575	0.51	3.2E-01	AJ277661.1	NT	Homo sapiens partial LMO1 gene for LIM domain only 1 protein, exon 1
8072	20766	33895	1.48	3.2E-01	M60266.1	NT	Rat ISO-atrial natriuretic factor gene, complete cds
8164	20858	33990	0.45	3.2E-01	AJ231001.1	NT	Rattus norvegicus repeat; map NOS-D12Wox1
8265	20959	34098	14.41	3.2E-01	X02508.1	NT	H.sapiens gene fragment for acetylcholine receptor (AChR) alpha subunit exons 8, 9 and 3' flanking region
8268	20962	34103	13.76	3.2E-01	BF311635.1	EST_HUMAN	601897107F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4126633 5'
8361	21054		1.38	3.2E-01	AL161574.2	NT	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 70

Table 4
Single Exon Probes Expressed in Brain

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
8398	21091	34226	1.24	3.2E-01	BF246771.1	EST_HUMAN	601855580F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4075627 5'
8398	21091	34227	1.24	3.2E-01	BF246771.1	EST_HUMAN	601855580F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4075627 5'
8471	21163	34306	2.65	3.2E-01	AE002015.1	NT	Deinococcus radiodurans R1 section 152 of 229 of the complete chromosome 1
8571	21263	34401	0.84	3.2E-01	U51026.1	NT	Oryctolagus cuniculus Ig H-chain pseudogene, V-region (VH6-a2) gene, partial cds
8571	21263	34402	0.84	3.2E-01	U51026.1	NT	Oryctolagus cuniculus Ig H-chain pseudogene, V-region (VH6-a2) gene, partial cds
8965	21656	34807	0.51	3.2E-01	AL163204.2	NT	Homo sapiens chromosome 21 segment HS21C004
8976	21666		2.18	3.2E-01	M86511.1	NT	Human monocyte antigen CD14 (CD14) mRNA, complete cds
9048	21737	34894	0.65	3.2E-01	AF041829.1	NT	Homo sapiens 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PF2K) gene, exons 12 and 13
9048	21737	34895	0.65	3.2E-01	AF041829.1	NT	Homo sapiens 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PF2K) gene, exons 12 and 13
9894	22544	35737	3.33	3.2E-01	U44914.1	NT	Borrelia burgdorferi plasmid cp32-2, erpC and erpD genes, complete cds; and unknown genes
10099	22747	35962	0.45	3.2E-01	BE326230.1	EST_HUMAN	hv99f05.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3181569 3'
10210	22858		3.41	3.2E-01	AB011399.1	NT	Homo sapiens gene for AF-6, complete cds
10566	23261	36498	3.94	3.2E-01	T06813.1	EST_HUMAN	EST04702 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA clone HFBDZ21
12010	25317		3.91	3.2E-01	L07288.1	NT	Drosophila melanogaster laminin A (Lam-A) mRNA, complete cds
12392	25374		1.44	3.2E-01	BE886846.1	EST_HUMAN	601507820F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3909532 5'
12524	24871		4.21	3.2E-01	O83217	SWISSPROT	ELONGATION FACTOR TU (EF-TU)
12655	24955		2.07	3.2E-01	L39874.1	NT	Homo sapiens deoxycytidylate deaminase gene, complete cds
12712	25354	30606	1.75	3.2E-01	BE385776.1	EST_HUMAN	601275480F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3616746 5'
2677	15386	28128	2.89	3.1E-01	R18051.1	EST_HUMAN	ye90h06.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:125051 5' similar to gb:M64241 QM PROTEIN (HUMAN);
2702	15532	28145	3.39	3.1E-01	7661971	NT	Homo sapiens KIAA0174 gene product (KIAA0174), mRNA
2702	15532	28146	3.39	3.1E-01	7661971	NT	Homo sapiens KIAA0174 gene product (KIAA0174), mRNA
2862	15630		1.29	3.1E-01	AW629036.1	EST_HUMAN	hi46h08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2975391 3'
3170	15933		3.35	3.1E-01	AB029069.1	NT	Mus musculus gene for Ser/Thr kinase KKIAMRE, exon 6
3887	16637	29276	0.8	3.1E-01	AJ251586.1	NT	Daucus carota mRNA for transcription factor E2F (E2F gene)
4908	17636	30250	0.73	3.1E-01	AE003984.1	NT	Xylella fastidiosa, section 130 of 229 of the complete genome
5390	18190	30882	9.73	3.1E-01	AF176111.1	NT	Homo sapiens hepatocyte nuclear factor-3 alpha (HNF3A) gene, exon 1
5513	18311	31212	0.73	3.1E-01	P44132	SWISSPROT	HYPOTHETICAL PROTEIN HI1236
5514	18312	31213	0.67	3.1E-01	Z74883.1	NT	S.cerevisiae chromosome XV reading frame ORF YOL141w
5524	18322		0.88	3.1E-01	Y13278.1	NT	Mus musculus mRNA for polycystin
5685	18478	31396	2.11	3.1E-01	AF184122.1	NT	Homo sapiens filamin 2 (FLN2) gene, exons 10 through 22
6191	25087	31942	0.59	3.1E-01	R94322.1	EST_HUMAN	yq41f04.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:198367 5'